

Result No.	Query			ID	Description
	Score	Match	Length		
1	2249	65.0	2404	US-08-311-023-3	Sequence 3, Appli
2	1288	37.2	1310	US-08-311-023-1	Sequence 1, Appli
3	360	10.4	385	US-09-712-016-78	Sequence 78, Appli
c 4	240	6.9	241	US-09-389-681-360	Sequence 360, App
c 5	240	6.9	241	US-09-620-405B-360	Sequence 360, App
c 6	240	6.9	241	US-09-433-826B-360	Sequence 360, App
c 7	240	6.9	241	US-09-604-287A-360	Sequence 360, App
c 8	165	4.8	177	US-09-020-956-162	Sequence 162, App
c 9	165	4.8	177	US-09-030-607-162	Sequence 162, App
c 10	165	4.8	177	US-09-439-313-162	Sequence 162, App
c 11	165	4.8	177	US-09-352-616A-162	Sequence 162, App
c 12	165	4.8	177	US-09-233-149A-162	Sequence 162, App
c 13	110.6	3.2	1888	US-09-599-360B-38	Sequence 38, Appli
c 14	59.8	1.7	285	US-08-990-571-64	Sequence 64, Appli
c 15	59.8	1.7	285	US-09-528-784A-64	Sequence 64, Appli
c 16	59.8	1.7	285	US-09-569-098A-64	Sequence 64, Appli
c 17	59.8	1.7	342	US-08-990-571-65	Sequence 65, Appli
c 18	59.8	1.7	342	US-09-528-784A-65	Sequence 65, Appli
c 19	59.8	1.7	342	US-09-569-098A-65	Sequence 65, Appli
c 20	59.8	1.7	351	US-08-990-571-60	Sequence 60, Appli
c 21	59.8	1.7	351	US-09-528-784A-60	Sequence 60, Appli
c 22	59.8	1.7	351	US-09-569-098A-60	Sequence 60, Appli
c 23	59.8	1.7	356	US-08-990-571-63	Sequence 63, Appli
c 24	59.8	1.7	356	US-09-528-784A-63	Sequence 63, Appli
c 25	59.8	1.7	356	US-09-569-098A-63	Sequence 63, Appli
c 26	59.8	1.7	358	US-08-990-571-58	Sequence 58, Appli
c 27	59.8	1.7	358	US-09-528-784A-58	Sequence 58, Appli

Dd	1	GGTTTATAGCAATTCATCATCAGTTTCCTGCTCTGCTGGGGTTATCTTAGTGCT	60
QY	1179	CTCARGAATCGGGTGTCTTTTCAAAATTCCTGAGTTTCCTTGTGSCACTGGCCGTTGGG	1238
Dd	61	CTCATGAATCGGGTGTCTTTTCAAAATTCCTGAGTTTCCTTGTGSCACTGGCCGTTGGG	120
QY	1239	ACTTTGAGTGTGATGCTTTTACACCTTCCTCCACATTCCTCATCAAGTCAACCAAT	1298
Dd	121	ACTTTGAGTGTGATGCTTTTACACCTTCCTCCACATTCCTCATCAAGTCAACCAAT	180
QY	1299	AGTCATAGCCATGAGAACAGCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT	1358
Dd	181	AGTCATAGCCATGAGAACAGCAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT	240
QY	1359	TCTTCTCAAACATAGAGAAAGTGCCTATTTTGAATTCCTCCAGTGGAGGCTCTACAGCT	1418
Dd	241	TCCTTCTCAAACATAGAGAAAGTGCCTATTTTGAATTCCTCCAGTGGAGGCTCTACAGCT	300
QY	1419	CTAGGAGGCTGTATTCATGTTCTTGTGTAACATGCTCTCAATTCATGATCAAAATTT	1478
Dd	301	CTAGGAGGCTGTATTCATGTTCTTGTGTAACATGCTCTCAATTCATGATCAAAATTT	360
QY	1479	AAAGATAGAGAAAAGAAATGAGAAACCTGAAATGAGATGAGATGAGATGAGATGAGAT	1538
Dd	361	AAAGATAGAGAAAAGAAATGAGAAACCTGAAATGAGATGAGATGAGATGAGATGAGAT	420
QY	1539	AAGCAGTCTCCAGTATGATCTCACTTTCACAAATGAGGAGAGTATGATACAGAT	1598
Dd	421	AAGCAGTCTCCAGTATGATCTCACTTTCACAAATGAGGAGAGTATGATACAGAT	480
QY	1599	GATCGAATGAGGCTATTTACGAGCAGACTCACAGAGCCCTCCACCTTTGATTCCTAG	1658
Dd	481	GATCGAATGAGGCTATTTACGAGCAGACTCACAGAGCCCTCCACCTTTGATTCCTAG	540
QY	1659	CAGCCTGAGTCTGGAAGAGAGAGGTCATGATGATGATGATGATGATGATGATGATGAT	1718
Dd	541	CAGCCTGAGTCTGGAAGAGAGAGGTCATGATGATGATGATGATGATGATGATGATGAT	600
QY	1719	TACAATGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1778
Dd	601	TACAATGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	660
QY	1779	CTGGCCGCTGAGAGATCTCATTCACCAACATGATGATGATGATGATGATGATGATGAT	1838
Dd	661	CTGGCCGCTGAGAGATCTCATTCACCAACATGATGATGATGATGATGATGATGATGAT	720
QY	1839	CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	1898
Dd	721	CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	780
QY	1899	AAAGATGCGGCGTGGCCATTTGGCCCTGGATGGTGAATATGGGTGATGCGCTGCACAT	1958
Dd	781	AAAGATGCGGCGTGGCCATTTGGCCCTGGATGGTGAATATGGGTGATGCGCTGCACAT	840
QY	1959	TTGAGGATGCGGCGTGGCCATTTGGCCCTGGATGGTGAATATGGGTGATGCGCTGCACAT	2018
Dd	841	TTGAGGATGCGGCGTGGCCATTTGGCCCTGGATGGTGAATATGGGTGATGCGCTGCACAT	900
QY	2019	ACTTCTGTTGCTGTCTTCTGTCATGATGATGATGATGATGATGATGATGATGATGAT	2078
Dd	901	ACTTCTGTTGCTGTCTTCTGTCATGATGATGATGATGATGATGATGATGATGATGAT	960
QY	2079	CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTATATGATGATGATGATGATGATGAT	2138
Dd	961	CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTATATGATGATGATGATGATGATGAT	1020
QY	2139	GCCTATCTGGATGCGACAGGAAATTTTCATGCTGCTATGCTGCTGCTGCTGCTGCTGCT	2198
Dd	1021	GCCTATCTGGATGCGACAGGAAATTTTCATGCTGCTATGCTGCTGCTGCTGCTGCTGCT	1080
QY	2199	TGGATATTTGCATCTACTGCTGGCTTATTCATGATGATGATGATGATGATGATGATGAT	2258
Dd	1081	TGGATATTTGCATCTACTGCTGGCTTATTCATGATGATGATGATGATGATGATGATGAT	1140
QY	2259	GAAATGCTGCACAAATGATGCTAGTGACCATGATGATGATGATGATGATGATGATGATGAT	2318
Dd	1141	GAAATGCTGCACAAATGATGCTAGTGACCATGATGATGATGATGATGATGATGATGATGAT	1200
QY	2319	CAGAATGCTGGATGCTTTTGGGTTTGGAAATATGTTTACTTATTTCCATATTTGAACAT	2378
Dd	1201	CAGAATGCTGGATGCTTTTGGGTTTGGAAATATGTTTACTTATTTCCATATTTGAACAT	1259
QY	2379	AAAAATGCTGGTTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTCTTAAAAAGT	2438
Dd	1260	AAAAATGCTGGTTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTCTTAAAAAGT	1318
QY	2439	TGTCATATTTTCAGTAGGTCATAGGAGATGAGTTTCTATGCTGCTGCTGCTGCTGCTGCTGCT	2498
Dd	1319	TGTCATATTTTCAGTAGGTCATAGGAGATGAGTTTCTATGCTGCTGCTGCTGCTGCTGCTGCT	1378
QY	2499	AAATTTAGTGGGTTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTCTTAAAAAGT	2558
Dd	1379	AAATTTAGTGGGTTTTCGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTCTTAAAAAGT	1438
QY	2559	AGTACGTTTAAATATTAAGTTTATCTTCTGCGATGATGATGATGATGATGATGATGATGAT	2618
Dd	1439	AGTACGTTTAAATATTAAGTTTATCTTCTGCGATGATGATGATGATGATGATGATGATGAT	1498
QY	2619	CGGTTATTACAGTTTATTTATGTAACAGAGATTTCGCGATGATGATGATGATGATGATGATGAT	2678
Dd	1499	CGGTTATTACAGTTTATTTATGTAACAGAGATTTCGCGATGATGATGATGATGATGATGATGAT	1558
QY	2679	AGGAAAAATGCTTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTT	2738
Dd	1559	AGGAAAAATGCTTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTT	1618
QY	2739	TAGGTTCTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2798
Dd	1619	TAGGTTCTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1678
QY	2799	AAAGAAAGCTTATCTGAATTTAAGCAAAAGAAATTAAGCAAAAGAAATTAAGCAAAAGAAATTAAG	2858
Dd	1679	AAAGAAAGCTTATCTGAATTTAAGCAAAAGAAATTAAGCAAAAGAAATTAAGCAAAAGAAATTAAG	1738
QY	2859	TTGGGAGGCGATAGATTCCTTATAAAATTCACAAATTTGTTGTTAAATTTAGAGGCGAGAA	2918
Dd	1739	TTGGGAGGCGATAGATTCCTTATAAAATTCACAAATTTGTTGTTAAATTTAGAGGCGAGAA	1798
QY	2919	TTTAAATTTAAGTATATAAAAGGAGCAATTTAGTATAGAGTACATTTAAACATTTTGT	2978
Dd	1799	TTTAAATTTAAGTATATAAAAGGAGCAATTTAGTATAGAGTACATTTAAACATTTTGT	1858
QY	2979	CAGGATTTATTTCCGTTAAAAACGCTAGTGAGCACTCTCATATCTATTAATTTAGTACATTTA	3038
Dd	1859	CAGGATTTATTTCCGTTAAAAACGCTAGTGAGCACTCTCATATCTATTAATTTAGTACATTTA	1918
QY	3039	ACTTTGTTAATACAGAAATCTAAATATATTTAAATGAAATTAAGCAAAATTAAGCAAAATTAAGCA	3098
Dd	1919	ACTTTGTTAATACAGAAATCTAAATATATTTAAATGAAATTAAGCAAAATTAAGCAAAATTAAGCA	1978
QY	3099	AGAAATTTGGAATTTCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3158
Dd	1979	AGAAATTTGGAATTTCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2038
QY	3159	TTTATGTTATCAGCAGACTGGGTTTATTTGCCAGTTATATATATATATATATATATATATATATAT	3218
Dd	2039	TTTATGTTATCAGCAGACTGGGTTTATTTGCCAGTTATATATATATATATATATATATATATATAT	2098
QY	3219	GATGTTCTGCTTACCTGGTTTACAAATTTATCAGAGTGT-----	3258
Dd	2099	GATGTTCTGCTTACCTGGTTTACAAATTTATCAGAGTGTAAACACTTTGATATATATGAG	2158
QY	3259	-----AAACCTTTGATATATGAG	3278
Dd	2159	GATATTAACAACTACACTAAGTATCATTTGATTCGATTCAGAAAACTTTGATATATATGAG	2218

```
QY 3279 GATATTAAACTACACTAAGTATCATTTGATTCGATTGAGAAAGTACTTTGATATCTC 3338
Db 2219 GATATTAAACTACACTAAGTATCATTTGATTCGATTGAGAAAGTACTTTGATATCTC 2278
QY 3339 AGTGCCTCAGTCTATCATTTGAGCAATGCTTTATATACGTTACTGTAGCCATACPA 3398
Db 2279 AGTGCCTCAGTCTATCATTTGAGCAATGCTTTATATACGTTACTGTAGCCATACPA 2338
QY 3399 GGCGCTGCTGGCAATCTCAGATGTTCTTTTACACAATAAATCCCTTATATACG 3458
Db 2339 GGCGCTGCTGGCAATCTCAGATGTTCTTTTACACAATAAATCCCTTATATACG 2398
QY 3459 TTG 3461
Db 2399 TTG 2401

RESULT 2
US-08-311-023-1
; Sequence 1, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; TITLE OF INVENTION: BREAST TUMOURS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32,925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1310 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1293
US-08-311-023-1

Query Match 37.2%; Score 1288; DB 1; Length 1310;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1310; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1119 GGTATTATAGCAATTCATCATCAGTTTCTCTCTCTGGGGTACTTACTAGTCCCT 1178
Db 1 GGTATTATAGCAATTCATCATCAGTTTCTCTCTCTGGGGTACTTACTAGTCCCT 60
QY 1179 CTCATGAATCGGGTGTGTTTCAAATTTCTCTGAGTTTCTTGTGGCACTGGCGGTGGG 1238
```

```
Db 61 CTCATGAATCGGGTGTGTTTCAAATTTCTCTGAGTTTCTTGTGGCACTGGCGGTGGG 120
QY 1239 ACTTTGAGTGGTGTGATGCTTTTACACCTTCTTCCACATTTCTCATGCAAGTACACACCAT 1298
Db 121 ACTTTGAGTGGTGTGATGCTTTTACACCTTCTTCCACATTTCTCATGCAAGTACACACCAT 180
QY 1299 AGTCATAGCCATGAGAAAGCAACGCAATGGAATGAAAGAGGACACCTTTTCAGTCACTG 1358
Db 181 AGTCATAGCCATGAGAAAGCAACGCAATGGAATGAAAGAGGACACCTTTTCAGTCACTG 240
QY 1359 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTATTCACGTTGGAGGGTCTACAGCT 1418
Db 241 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTATTCACGTTGGAGGGTCTACAGCT 300
QY 1419 CTAGGAGGCTGTATTTTCATGTTTCTTGTGAACTGTCTCTCACATTCATCAAACTTT 1478
Db 301 CTAGGAGGCTGTATTTTCATGTTTCTTGTGAACTGTCTCTCACATTCATCAAACTTT 360
QY 1479 AAAGATAAGAGAAAGAAAGATFCAGAAAGAACTGAAATGATGATGTGGAGATTAAG 1538
Db 361 AAAGATAAGAGAAAGAAAGATFCAGAAAGAACTGAAATGATGATGTGGAGATTAAG 420
QY 1539 AAGCAGTTCTCCAGTATGATTCCTCAACTTTCACAAATGAGGAGAAAGTACAGAT 1598
Db 421 AAGCAGTTCTCCAGTATGATTCCTCAACTTTCACAAATGAGGAGAAAGTACAGAT 480
QY 1599 GATCGAACTGGAAGGCTATTTACGAGCAGACTCACAGAGCCCTCCCACTTTGATTCAG 1658
Db 481 GATCGAACTGGAAGGCTATTTACGAGCAGACTCACAGAGCCCTCCCACTTTGATTCAG 540
QY 1659 CAGCCTGCACTTGGAAAGAGAGAGGTCATGATGATGATGATGATGATGATGATGATGATG 1718
Db 541 CAGCCTGCACTTGGAAAGAGAGAGGTCATGATGATGATGATGATGATGATGATGATGATG 600
QY 1719 TACAATGAATATGTACCCAGAGGGTGCAAGATAATGCAATTCACATTTCCACAGTACA 1778
Db 601 TACAATGAATATGTACCCAGAGGGTGCAAGATAATGCAATTCACATTTCCACAGTACA 660
QY 1779 CTCGGCCAGTCAGACGATCTCAATTCACACCATCATGACTACCATCATATTTCTCCATCAT 1838
Db 661 CTCGGCCAGTCAGACGATCTCAATTCACACCATCATGACTACCATCATATTTCTCCATCAT 720
QY 1839 CACCACCAACCAACCACTCTCACATTCACAGCCAGCGCTACTCTCGGGAGGAGCTG 1898
Db 721 CACCACCAACCAACCACTCTCACATTCACAGCCAGCGCTACTCTCGGGAGGAGCTG 780
QY 1899 AAAGATGCCGGCTGCCACTTTTGGCTGGATGGTGAATAATGGGTGATGGCTGCACAA 1958
Db 781 AAAGATGCCGGCTGCCACTTTTGGCTGGATGGTGAATAATGGGTGATGGCTGCACAA 840
QY 1959 TTCAGCGATGGCTTAGCAATTTGGTGTGCTTTTACTGAGGCTTATCAAGTGGTTTAACT 2018
Db 841 TTCAGCGATGGCTTAGCAATTTGGTGTGCTTTTACTGAGGCTTATCAAGTGGTTTAACT 900
QY 2019 ACTTCTGTTGCTGCTTCTGTCATGAGTGTGCTTCATGATAGTGTGCTTGTGCTTCA 2078
Db 901 ACTTCTGTTGCTGCTTCTGTCATGAGTGTGCTTCATGATAGTGTGCTTGTGCTTCA 960
QY 2079 CTAAAGGCTGGCATGACCGTTTAAAGCGGTGCTTTTAAATGCAATGTACAGCATGCTG 2138
Db 961 CTAAAGGCTGGCATGACCGTTTAAAGCGGTGCTTTTAAATGCAATGTACAGCATGCTG 1020
QY 2139 GCCTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTTCATTTGCTGAAATGTTTCTATG 2198
Db 1021 GCCTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTTCATTTGCTGAAATGTTTCTATG 1080
QY 2199 TGGATATTTGCACTTACTGCTGGCTTATTCATGATGTTGCTGCTGTTGATGTTGTTTACT 2258
Db 1081 TGGATATTTGCACTTACTGCTGGCTTATTCATGATGTTGCTGCTGTTGATGTTGTTTACT 1140
QY 2259 GAAATGCTGCACATGATGCTAGTGCACATGATGATGATGATGATGATGATGATGATGATG 2318
```

Db 1141 GAAATGCTGCACAAATGATGCTAGTGAACATGATGATGAGCGCTGGGGTATTTCTTTTAA 1200
QY 2319 CAGATGCTGGGATGCTTTGGGTTTGGGAATTAATGTTACTATTTCCATATTTGAACAT 2378
Db 1201 CAGATGCTGGGATGCTTTGGGTTTGGGAATTAATGTTACTATTTCCATATTTGAACAT 1259
QY 2379 AAATGCTGTTTGGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTAGCT 2430
Db 1260 AAATGCTG-TTGGTATAAATTTCTAGTTAAAGTTTAAATGCTAGAGTAGCT 1310

RESULT 3

US-09-712-016-78
; Sequence 78, Application US/09712016
; Patent No. 6468790
; GENERAL INFORMATION:
; APPLICANT: Giese, Klaus
; TITLE OF INVENTION: METASTATIC BREAST AND COLON CANCER
; FILE REFERENCE: 200130.460
; CURRENT APPLICATION NUMBER: US/09/712,016
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US/09/417,615
; PRIOR FILING DATE: PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 78
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-712-016-78

Query Match 10.4%; Score 360; DB 4; Length 385;
Best Local Similarity 98.2%; Pred. No. 1.3e-90;
Matches 375; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 2453 TAGGTCATPAGGAGATGAGTGTGATGCTGATCTATCCAGCGTTAAAGTTAGTGGGTTT 2512
Db 1 TCGGTCATPAGGAGAA--GATTTGATGCTGATCTATCCAGCGTTAAAGTTAGTGGGTTT 58
QY 2513 TGTGATTTTCTATGATTAATGCTGCTGTGTTACAAAGTCAGTTAAAGTACGTTTAAAT 2572
Db 59 TGTGATTTTCTATGATTAATGCTGCTGTGTTACAAAGTCAGTTAAAGTACGTTTAAAT 118
QY 2573 ATTTAAGTTATTTCTATGCGAGATAAAATCTGATGCAATTCACCGGTTATACCAAGT 2632
Db 119 ATTTAAGTTATTTCTATGCGAGATAAAATCTGATGCAATTCACCGGTTATACCAAGT 178
QY 2633 TTATATGCTAAACAGAGATTTGGCATGACATGTTCTGATGTTTCAGGGAAAAATGCT 2592
Db 179 TTATATGCTAAACAGAGATTTGGCATGACATGTTCTGATGTTTCAGGGAAAAATGCT 238
QY 2693 TTAATGCTTTTCAAGAACTAACACAGTTATTCCTATCTAGTTAGGTTCTGGA 2752
Db 239 TTAATGCTTTTCAAGAACTAACACAGTTATTCCTATCTAGTTAGGTTCTGGA 298
QY 2753 ACTGCTGGTGTAGGAATAAGATGTCATGCAAGCCTAAATPACCAAGAAAGCTTATAC 2812
Db 299 ACTGCTGGTGTAGGAATAAGATGTCATGCAAGCCTAAATPACCAAGAAAGCTTATAC 358
QY 2813 TGAATTTAAGCAAGAAATAAA 2834
Db 359 TGAATTTAAGCAAGAAATAAA 380

RESULT 4

US-09-389-681-360/c
; Sequence 360, Application US/09389681a
; Patent No. 6518237
; GENERAL INFORMATION:
; APPLICANT: Yuqui, Jiang
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681a
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 360
; LENGTH: 241
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(241)
; OTHER INFORMATION: n = A,T,C or G
US-09-389-681-360

Query Match 6.9%; Score 240; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2720 TTATTCCTATPACTGGATTTTAGTCTCTGAAGAACTGCTGTTTAGAATAAGATGT 2779
Db 241 TTATTCCTATPACTGGATTTTAGTCTCTGAAGAACTGCTGTTTAGAATAAGATGT 182
QY 2780 GCATGAAGCCTAAATPACCAAGAAAGCTTATCTAGTAATTAAGCAAGAAATAAGGAGA 2839
Db 181 GCATGAAGCCTAAATPACCAAGAAAGCTTATCTAGTAATTAAGCAAGAAATAAGGAGA 122
QY 2840 AAAGAGAAAGAACTGAGAAATTTGGGAGGAGCATAGATTTTATAAAATCAGAAATTTGTT 2899
Db 121 AAAGAGAAAGAACTGAGAAATTTGGGAGGAGCATAGATTTTATAAAATCAGAAATTTGTT 62
QY 2900 GTAATTAAGAGGGAGAAATTTAGAAATTAAGTATAAAGGAGAAATTAAGTATAGTAC 2959
Db 61 GTAATTAAGAGGGAGAAATTTAGAAATTAAGTATAAAGGAGAAATTAAGTATAGTATAGTAC 2

RESULT 5

US-09-620-405B-360/C
; Sequence 360, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 360
; LENGTH: 241
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(241)
; OTHER INFORMATION: n = A,T,C or G
US-09-620-405B-360

Query Match 6.9%; Score 240; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2720 TTATTCCTATPACTGGATTTTAGTCTCTGAAGAACTGCTGTTTAGAATAAGATGT 2779
Db 241 TTATTCCTATPACTGGATTTTAGTCTCTGAAGAACTGCTGTTTAGAATAAGATGT 182

Db 241 TTATCTCTACTGGATTTTAGTCTCTGAGAACTGCTGGTGTCTTAGGAATAGATGT 182
QY 2780 GCATGAAGCTTAAATACCAAGAAGCTTATAGTGAATTTAAGCAAAAGAAATAAGGAGA 2839
Db 181 GCATGAAGCTTAAATACCAAGAAGCTTATAGTGAATTTAAGCAAAAGAAATAAGGAGA 122
QY 2840 AAGAGAGAAGTCTGAGAAATGGGAGGCATAGATCTTATAAAAATCACAATAATTTGT 2899
Db 121 AAGAGAGAAGTCTGAGAAATGGGAGGCATAGATCTTATAAAAATCACAATAATTTGT 62
QY 2900 GTAAATAGAGGGGAGAAATTTAGAAATTAAGTATAAAAGGAGCAATAGTATAGAGTAC 2959
Db 61 GTAAATAGAGGGGAGAAATTTAGAAATTAAGTATAAAAGGAGCAATAGTATAGAGTAC 2

RESULT 6

US-09-433-826B-360/c
; Sequence 360, Application US/09433826B
; Patent No. 6579973

GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui

; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C4

; CURRENT APPLICATION NUMBER: US/09/433,826B
; CURRENT FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 474
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 360
; LENGTH: 241

; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(241)
; OTHER INFORMATION: n = A,T,C or G

US-09-433-826B-360

Query Match 6.9%; Score 240; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;

Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2720 TTATCTCTACTGGATTTTAGTCTCTGAGAACTGCTGGTGTCTTAGGAATAGATGT 2779

Db 241 TTATCTCTACTGGATTTTAGTCTCTGAGAACTGCTGGTGTCTTAGGAATAGATGT 182

QY 2780 GCATGAAGCTTAAATACCAAGAAGCTTATAGTGAATTTAAGCAAAAGAAATAAGGAGA 2839

Db 181 GCATGAAGCTTAAATACCAAGAAGCTTATAGTGAATTTAAGCAAAAGAAATAAGGAGA 122

QY 2840 AAGAGAGAAGTCTGAGAAATGGGAGGCATAGATCTTATAAAAATCACAATAATTTGT 2899

Db 121 AAGAGAGAAGTCTGAGAAATGGGAGGCATAGATCTTATAAAAATCACAATAATTTGT 62

QY 2900 GTAAATAGAGGGGAGAAATTTAGAAATTAAGTATAAAAGGAGCAATAGTATAGAGTAC 2959

Db 61 GTAAATAGAGGGGAGAAATTTAGAAATTAAGTATAAAAGGAGCAATAGTATAGAGTAC 2

RESULT 7

US-09-604-287A-360/c

; Sequence 360, Application US/09604287A
; Patent No. 6586572

GENERAL INFORMATION:

; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.

; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C7

; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22

; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 360
; LENGTH: 241

; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(241)
; OTHER INFORMATION: n = A,T,C or G

US-09-604-287A-360

Query Match 6.9%; Score 240; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;

Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2720 TTATCTCTACTGGATTTTAGTCTCTGAGAACTGCTGGTGTCTTAGGAATAGATGT 2779

Db 241 TTATCTCTACTGGATTTTAGTCTCTGAGAACTGCTGGTGTCTTAGGAATAGATGT 182

QY 2780 GCATGAAGCTTAAATACCAAGAAGCTTATAGTGAATTTAAGCAAAAGAAATAAGGAGA 2839

Db 181 GCATGAAGCTTAAATACCAAGAAGCTTATAGTGAATTTAAGCAAAAGAAATAAGGAGA 122

QY 2840 AAGAGAGAAGTCTGAGAAATGGGAGGCATAGATCTTATAAAAATCACAATAATTTGT 2899

Db 121 AAGAGAGAAGTCTGAGAAATGGGAGGCATAGATCTTATAAAAATCACAATAATTTGT 62

QY 2900 GTAAATAGAGGGGAGAAATTTAGAAATTAAGTATAAAAGGAGCAATAGTATAGAGTAC 2959

Db 61 GTAAATAGAGGGGAGAAATTTAGAAATTAAGTATAAAAGGAGCAATAGTATAGAGTAC 2

RESULT 8

US-09-020-956-162/c
; Sequence 162, Application US/09020956
; Patent No. 6261562

GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHOD;

; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle
; STATE: WA

; COUNTRY: USA
; ZIP: 98104

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 162:
; SEQUENCE CHARACTERISTICS:


```
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 162
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-352-616A-162

Query Match
Best Local Similarity 4.8%; Score 165; DB 4; Length 177;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3159 TTATGTATACACAGCTGGTTATGTCACAGTATATATACACAAAGCTGTATGACTG 3218
DB 165 TTATGTATACACAGCTGGTTATGTCACAGTATATATACACAAAGCTGTATGACTG 106

QY 3219 GATGTTCTGGTTACCTGGTTTACAAAATATACAGTAGTAAACTTTTGATATATAGAG 3278
DB 105 GATGTTCTGGTTACCTGGTTTACAAAATATACAGTAGTAAACTTTTGATATATAGAG 46

QY 3279 GATATTAACACACTAGTATATATATATATATATATATATATATATATATATATAT 3323
DB 45 GATATTAACACACTAGTATATATATATATATATATATATATATATATATATATATAT 1

RESULT 12
US-09-232-149A-162/c
; Sequence 162, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 162
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-232-149A-162

Query Match
Best Local Similarity 4.8%; Score 165; DB 4; Length 177;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3159 TTATGTATACACAGCTGGTTATGTCACAGTATATATATACACAAAGCTGTATGACTG 3218
DB 165 TTATGTATACACAGCTGGTTATGTCACAGTATATATATATACACAAAGCTGTATGACTG 106

QY 3219 GATGTTCTGGTTACCTGGTTTACAAAATATACAGTAGTAAACTTTTGATATATAGAG 3278
DB 105 GATGTTCTGGTTACCTGGTTTACAAAATATACAGTAGTAAACTTTTGATATATAGAG 46

QY 3279 GATATTAACACACTAGTATATATATATATATATATATATATATATATATATATAT 3323
DB 45 GATATTAACACACTAGTATATATATATATATATATATATATATATATATATATATAT 1
```

```
RESULT 13
US-09-599-360B-38
; Sequence 38, Application US/09599360B
; Patent No. 6548633
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Bougueleret, L.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
```

```
; FILE REFERENCE: GENSET.050CP3
; CURRENT APPLICATION NUMBER: US/09/599,360B
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 60/113,686
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/141,032
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 09/469,099
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: Patent.pm
; SEQ ID NO 38
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 139..1389
; NAME/KEY: sig_peptide
; LOCATION: 139..198
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.00
; OTHER INFORMATION: seq HLLAGFCVWVLG/VW
; NAME/KEY: polyA_signal
; LOCATION: 1854..1859
; NAME/KEY: polyA_site
; LOCATION: 1873..1888
US-09-599-360B-38

Query Match
Best Local Similarity 3.2%; Score 110.6; DB 4; Length 1888;
Matches 211; Conservative 0; Mismatches 149; Indels 1; Gaps 1;

QY 1908 GGCCTCGCCACTTTGGCCCTGGATGATGAATGAGTGGTGGCTGCACCAATTCAGCGAT 1967
DB 1273 GGTGGCACTGATATACAGTGGATGGTCTCTCTGGAGATGGTCTACACAACTCACTGAT 1332

QY 1968 GGCCTAGCAATTTGGTGGCTTTTACTCAAGCTTATCAAGTGGTTAAGTACTTCTGTT 2027
DB 1333 GGGCTGGCCATAGGTGCTGCTCTCTCTGATGGCTTCTC-CGGGGCTCAGTACACCTTA 1391

QY 2028 GCTGTGTTCTGTCATGAGTTGCTCATGAATTAGTGGTGGTGGTGGTGGTGGTGGTGGT 2087
DB 1392 GCGGCTCTCTGCCATGAGCTGCCACCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1451

QY 2088 GGCATGACCGTTAAGCAGGCTGTCCTTTATATATGATGATGATGATGATGATGATGATG 2147
DB 1452 GGGCTGCTCTTTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1511

QY 2148 GGAATGGCAACAGGAATTTTCATTTGCTGATGATGATGATGATGATGATGATGATGAT 2207
DB 1512 GGTGCACTGCTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGG 1571

QY 2208 GCACCTACTGCTGGCTTTATTCATGATGATGATGATGATGATGATGATGATGATGATG 2267
DB 1572 GGGGCTACTGCTGGGGCTTCTCTCTATGTTGGCCCTTGTGGACATGCTACACGCCCTG 1631

QY 2268 C 2268
DB 1632 C 1632
```

```
RESULT 14
US-08-990-571-64/c
; Sequence 64, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
```

QY 405 ATCCATATACCACTGACCAACGACCATCACTCAGACCAAGCATCACTCAGACCATGAG 464

1	3336	96.4	3586	14	US-10-176-847-51	Sequence 51, Appl
2	3335.6	96.4	3537	14	US-10-198-846-10007	Sequence 10007, A
3	2716	78.5	2744	14	US-10-171-581-170	Sequence 170, App
4	2716	78.5	2744	14	US-10-177-293-263	Sequence 263, App
5	1141.4	36.0	1193	10	US-09-925-300-591	Sequence 591, App
6	900.4	26.0	1449	14	US-10-198-846-10321	Sequence 10321, A
7	694.2	20.1	1032	14	US-10-198-846-13805	Sequence 13805, A
8	578.4	16.7	606	14	US-10-198-846-8490	Sequence 8490, Ap
9	578.4	16.7	1598	14	US-10-198-846-11215	Sequence 11215, A
10	404.8	11.7	483	10	US-09-867-701-2021	Sequence 2021, Ap
11	390.8	11.3	395	9	US-09-910-689-94	Sequence 94, Appl
12	390.8	11.3	395	9	US-09-910-689-94	Sequence 94, Appl
13	390.8	11.3	395	13	US-10-010-742-94	Sequence 94, Appl
14	389.6	11.3	504	11	US-09-918-995-22272	Sequence 22272, A
15	360	10.4	385	9	US-09-827-669-78	Sequence 78, Appl
16	360	10.4	385	12	US-10-236-006-78	Sequence 78, Appl

—

100

Db 2505 TTCTAGTTAGGTTAAAGCTAGAGTAGCTTAAAAAGTTGTCATAGTTTCCAGTAGGTCA 2564
Qy 2460 TAGGAGATGATGTTGTATGCTGTACTATGACAGCGTTTAAAGTTAGTGGTTTGTGATT 2519
Db 2565 TAGGAGATGATGTTGTATGCTGTACTATGACAGCGTTTAAAGTTAGTGGTTTGTGATT 2624
Qy 2520 TTGTGATTAATGATGCTGTCTGTACAAAGTCAGTTTAAAGGTACGTTTAAATTAATTAAG 2579
Db 2625 TTGTGATTAATGATGCTGTCTGTACAAAGTCAGTTTAAAGGTACGTTTAAATTAATTAAG 2684
Qy 2580 TTATTCATCTCTGGAGATAAATCTGTATGTGCAATTCACGGTATTACCAAGTTTATTTAT 2639
Db 2685 TTATTCATCTCTGGAGATAAATCTGTATGTGCAATTCACGGTATTACCAAGTTTATTTAT 2744
Qy 2640 GTAACAAGATTTGGCATGACATGTTCTGTATGTTTTCAGGAAAAATGCTTTTAATGTC 2699
Db 2745 GTAACAAGATTTGGCATGACATGTTCTGTATGTTTTCAGGAAAAATGCTTTTAATGTC 2804
Qy 2700 TTTTTCAGAACTAACACAGTATTCCTATATCTGATTTAGCTCTCTGAGAGCTGCTG 2759
Db 2805 TTTTTCAGAACTAACACAGTATTCCTATATCTGATTTAGCTCTCTGAGAGCTGCTG 2864
Qy 2760 GRTTGTAGAAATAAGATGTGATGAGCCTAAATACCAAGAAAGCTTTATCTGAATTT 2819
Db 2865 GRTTGTAGAAATAAGATGTGATGAGCCTAAATACCAAGAAAGCTTTATCTGAATTT 2924
Qy 2820 AAGCAAAAGAAATAAGGAGAAAGAGAGATCTGAGAAATGGGAGGAGATAGATCTTA 2879
Db 2925 AAGCAAAAGAAATAAGGAGAAAGAGAGATCTGAGAAATGGGAGGAGATAGATCTTA 2984
Qy 2880 TAAATACACAAAATTTGTTGTAATTAGAGGGAGAAATTTAGAATTAAGTATTAATAAG 2939
Db 2985 TAAATACACAAAATTTGTTGTAATTAGAGGGAGAAATTTAGAATTAAGTATTAATAAG 3044
Qy 2940 GCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2999
Db 3045 GCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3104
Qy 3000 CGTAGTGAGCACTCTCATATACATTAATGATGATGATGATGATGATGATGATGATGATG 3059
Db 3105 CGTAGTGAGCACTCTCATATACATTAATGATGATGATGATGATGATGATGATGATGATG 3164
Qy 3060 TAAATATATTTAATGAATTAAGCAATATACACTTGACCAAGAAATTTGGAATTTCAAAAT 3119
Db 3165 TAAATATATTTAATGAATTAAGCAATATACACTTGACCAAGAAATTTGGAATTTCAAAAT 3224
Qy 3120 GTTCGTGGGGTTATATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3179
Db 3225 GTTCGTGGGGTTATATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3284
Qy 3180 TTATTGCCAACTTATATATACCAAAAGCTGTATGATGATGATGATGATGATGATGATGATG 3239
Db 3285 TTATTGCCAACTTATATATACCAAAAGCTGTATGATGATGATGATGATGATGATGATGATG 3344
Qy 3240 ACAAAATATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3299
Db 3345 ACAAAATATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3404
Qy 3300 ATCATTTGATTCGATTCAGAAAGTACTTTGATATCTCTCAGTGTCTCAGTGTCTCAGTGT 3359
Db 3405 ATCATTTGATTCGATTCAGAAAGTACTTTGATATCTCTCAGTGTCTCAGTGTCTCAGTGT 3464
Qy 3360 TGAGCAATTTGCTTTATATACGGTACTGTAGCCATFAGTGGCTGCTGTGTCATCTCTCT 3419
Db 3465 TGAGCAATTTGCTTTATATACGGTACTGTAGCCATFAGTGGCTGCTGTGTCATCTCTCT 3524
Qy 3420 AGATGTTCTCTTTTACACAAATAAATTCCTTATATFACGCTTG 3461
Db 3525 AGATGTTCTCTTTTACACAAATAAATTCCTTATATFACGCTTG 3566

US-10-198-846-10007
; Sequence 10007, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10007
; LENGTH: 3537
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3533, 3534, 3535, 3536, 3537
; OTHER INFORMATION: n = A, T, C or G
US-10-198-846-10007

Query Match 96.44; Score 3335.6; DB 14; Length 3537;
Best Local Similarity 99.24; Pred. No. 0;
Matches 3433; Conservative 0; Mismatches 4; Indels 25; Gaps 7;

Qy 3 CGTCCGGAATTCGCGACGAGCGCGTTCGCGCCCTGGTAGAGATTTCTCGAAGACACC 62
Db 3 CGTCCGGAATTCGCGACGAGCGCGTTCGCGCCCTGGTAGAGATTTCTCGAAGACACC 62
Qy 63 AGTGGGCGCGTGTGGACCAACCTGCGCGGTGGCGGCGCGCGCGCGCGCGCGCGCG 122
Db 63 AGTGGGCGCGTGTGGACCAACCTGCGCGGTGGCGGCGCGCGCGCGCGCGCGCGCG 122
Qy 123 GGAGACGAGCGCGCAATGCGGAGGAGTATCTGTATCTTGATCTGACCTTGGCCCTC 182
Db 123 GGAGACGAGCGCGCAATGCGGAGGAGTATCTGTATCTTGATCTGACCTTGGCCCTC 182
Qy 183 TCTGTACAAATCCCTCATGAACTTAAAGCAGCTCTTCCCGCAGACCACTGAGAAA 242
Db 183 TCTGTACAAATCCCTCATGAACTTAAAGCAGCTCTTCCCGCAGACCACTGAGAAA 242
Qy 243 ATTAGTCGGAATTTGGGAATCTGGCAATTAATGTTGACCTTGGCAATTTCCACAGGCAATAT 302
Db 243 ATTAGTCGGAATTTGGGAATCTGGCAATTAATGTTGACCTTGGCAATTTCCACAGGCAATAT 302
Qy 303 CATCTACACAGCTTTTCTACCGCTATGAGAAATTAATCTTTGTCAGTTCAAGGGTTC 362
Db 303 CATCTACACAGCTTTTCTACCGCTATGAGAAATTAATCTTTGTCAGTTCAAGGGTTC 362
Qy 363 AGAAAATTTACTTCAAAATATAGCATAGATTAAGATTAAAAAGATTCCATATACACCATGAC 422
Db 363 AGAAAATTTACTTCAAAATATAGCATAGATTAAGATTAAAAAGATTCCATATACACCATGAC 422
Qy 423 CAGGACCATCTCAGACACGAGCATCAGTACAGACCATGAGCGTCACTCAGACCATGAG 482
Db 423 CAGGACCATCTCAGACACGAGCATCAGTACAGACCATGAGCGTCACTCAGACCATGAG 482
Qy 483 CATCACTCAGACACGAGCATCAGTACCATCATCATCTCTCACCATAAATCATGCT 542
Db 483 CATCACTCAGACACGAGCATCAGTACCATCATCATCTCTCACCATAAATCATGCT 542
Qy 543 GCTTCTGTTAAAAATAAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTCAAGT 602
Db 525 GCTTCTGTTAAAAATAAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTCAAGT 584
Qy 603 AAAGATCTTAGAAACAGCCAGGGGAAAGAGCTCACCAGACCAATGCCAGTGGTAGA 662

D0	585	AAAGATCTTAGAAACAGCCAGGGGAAAGAGCTCACCGACCGAACAATCCCAAGTGGTAGA	644
QY	663	AGGAATGTCAGAGCAGCTGTAGTGTCTAGTGAAGTGACCTCAACTGTGTACAAACACTGTC	722
D0	645	AGGAATGTCAGAGCAGCTGTAGTGTCTAGTGAAGTGACCTCAACTGTGTACAAACACTGTC	704
QY	723	TCTGAGGAACTCACTTCTTAGAGACAATAGAGACTCCAGAGCCTGGAAAACTCTTCCCC	782
D0	705	TCTGAGGAACTCACTTCTTAGAGACAATAGAGACTCCAGAGCCTGGAAAACTCTTCCCC	764
QY	783	AAAGATGTAAGCAGCTCCACTCCACCAGTGTCAATCAAAGAGCGGGTGAGCGCGCTG	842
D0	765	AAAGATGTAAGCAGCTCCACTCCACCAGTGTCAATCAAAGAGCGGGTGAGCGCGCTG	824
QY	843	GCTGGTAGGAAACAAATGAATCTGTGAGTGAGCCCGCGAAAAAGGCTTTATGTATTCCAGA	902
D0	825	GCTGGTAGGAAACAAATGAATCTGTGAGTGAGCCCGCGAAAAAGGCTTTATGTATTCCAGA	884
QY	903	AACCAAAATGAATAATCTCTAGAGAGTGTTCATATGCATCAAGATCAAGTACATGACATCTCATGGC	962
D0	885	AACCAAAATGAATAATCTCTAGAGAGTGTTCATATGCATCAAGTACATGACATCTCATGGC	944
QY	963	ATGGGCAATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTGTGCCAGCCATCATC	1022
D0	945	ATGGGCAATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTGTGCCAGCCATCATC	1004
QY	1023	AAACCAATGAGCTAGATCTGTCTCATCTCATCAAGTGGAAGAGAGCTGAATCCCT	1082
D0	1005	AAACCAATGAGCTAGATCTGTCTCATCTCATCAAGTGGAAGAGAGCTGAATCCCT	1064
QY	1083	CCAAAGACCTATTCAATTAACAATAGCCTGGGTGGTGGTTTATAGCCATTTCCCATCATC	1142
D0	1065	CCAAAGACCTATTCAATTAACAATAGCCTGGGTGGTGGTTTATAGCCATTTCCCATCATC	1124
QY	1143	AGTTTCTCTGCTCTGCTGGGGTTATCTTAGTGCCCTCTCATGAATCGGGTGTGTTTCAA	1202
D0	1125	AGTTTCTCTGCTCTGCTGGGGTTATCTTAGTGCCCTCTCATGAATCGGGTGTGTTTCAA	1184
QY	1203	TTTCTCTCTGAGTTTCTTGTGGCACTGGCGCTGGGACTTTGAGTGGTGATGCTTTTATA	1262
D0	1185	TTTCTCTCTGAGTTTCTTGTGGCACTGGCGCTGGGACTTTGAGTGGTGATGCTTTTATA	1244
QY	1263	CACCTCTTCCCACTTCTATGCAAGTCAACCACCATAGTATAGCCATGAAGAACCAGCA	1322
D0	1245	CACCTCTTCCCACTTCTATGCAAGTCAACCACCATAGTATAGCCATGAAGAACCAGCA	1304
QY	1323	ATGGAATGAAGAAGAGGACCACTTTTCAGTCACTGTCTCTCAAAACATAGCAAAAGT	1382
D0	1305	ATGGAATGAAGAAGAGGACCACTTTTCAGTCACTGTCTCTCAAAACATAGCAAAAGT	1364
QY	1383	GCCTATTGTTGATTCACCTGGAGGGTCTAACAGCTCTAGGAGCGCTGTATTTCATGTTT	1442
D0	1365	GCCTATTGTTGATTCACCTGGAGGGTCTAACAGCTCTAGGAGCGCTGTATTTCATGTTT	1424
QY	1443	CTGTGTGAACATGCTCTCACATTTGATCAACAAATTTAAGATAGAGAAAAGATACAG	1502
D0	1425	CTGTGTGAACATGCTCTCACATTTGATCAACAAATTTAAGATAGAGAAAAGATACAG	1484
QY	1503	AGAAGAACTGAAAATGATGATGATGTTGGAGATTAAAGACAGTTGTCCAAGTATGAATCT	1562
D0	1485	AGAAGAACTGAAAATGATGATGATGTTGGAGATTAAAGACAGTTGTCCAAGTATGAATCT	1544
QY	1563	CAACTTTCAACAAATGAGAGAAAGTAGATACAGATGATCGAACTGAAGGCTATTTTAGCA	1622
D0	1545	CAACTTTCAACAAATGAGAGAAAGTAGATACAGATGATCGAACTGAAGGCTATTTTAGCA	1604
QY	1623	GCAGACTCAAGAGGCGCTCCCACTTTGATTTCTCAGAGCGCTGCAGTCTTGGAAAGAGAA	1682
D0	1605	GCAGACTCAAGAGGCGCTCCCACTTTGATTTCTCAGAGCGCTGCAGTCTTGGAAAGAGAA	1664
QY	1683	GAGGTCTATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAGGG	1742

Db	1665	GAGGTCATGATAGCTCATGCTACTCCACAGGAAGTCTCAATGAATATGTATACCAGAGGG	1724
QY	1743	TGCAAGAATAAATGCCATTTCACATTTCCACAGATACACTCGGCCAGTCAGACGATCTCATTT	1802
Db	1725	TGCAAGAATAAATGCCATTTCACATTTCCACAGATACACTCGGCCAGTCAGACGATCTCATTT	1784
QY	1803	CACACACATCATGACTACCATCATCATATTCCTCCATCATCACACACACCAAAACACCATCTCT	1862
Db	1785	CACACACATCATGACTACCATCATCATATTCCTCCATCATCACACACCAAAACACCATCTCT	1844
QY	1863	CACAGTCACACGACGCGCTACTCTCGGAGAGAGCTGAAAGATCGCGCGTCGCGCACTTTG	1922
Db	1845	CACAGTCACACGACGCGCTACTCTCGGAGAGAGCTGAAAGATCGCGCGTCGCGCACTTTG	1904
QY	1923	GCGTGGATGGTGATTAAGGGTAGGGCTGCACAAATTCACAGATGGCGCTAGCAATTTGTT	1982
Db	1905	GCGTGGATGGTGATTAAGGGTAGGGCTGCACAAATTCACAGATGGCGCTAGCAATTTGTT	1964
QY	1983	GCTGCTTTTACTCAAGGCTTATCAAGTGGTTTAAGTACTTCTGTTGCTGTGTTCTGCTCAT	2042
Db	1965	GCTGCTTTTACTCAAGGCTTATCAAGTGGTTTAAGTACTTCTGTTGCTGTGTTCTGCTCAT	2024
QY	2043	GAGTTGCGCTCATGAATTAGGTGACTTTTGCTGTTCTACTAAAGGTGGCATGACCGTTTAAG	2102
Db	2025	GAGTTGCGCTCATGAATTAGGTGACTTTTGCTGTTCTACTAAAGGTGGCATGACCGTTTAAG	2084
QY	2103	CAGGCTGCTCTTATAATGATGTGCAGCCATCCTCGCGGTATCTTGGAAATGCGCAACAGGA	2162
Db	2085	CAGGCTGCTCTTATAATGATGTGCAGCCATCCTCGCGGTATCTTGGAAATGCGCAACAGGA	2144
QY	2163	ATTTTCATTTGGTCATTAATGCTGAAAAATGTTTCTATGTGGATATTTGCACTTACTGCTGCG	2222
Db	2145	ATTTTCATTTGGTCATTAATGCTGAAAAATGTTTCTATGTGGATATTTGCACTTACTGCTGCG	2204
QY	2223	TTATTCATGATGTTGCTCTGCTGTGATATGGTACCTGGAATGCTGCACATATGATGCTAGT	2282
Db	2205	TTATTCATGATGTTGCTCTGCTGTGATATGGTACCTGGAATGCTGCACATATGATGCTAGT	2264
QY	2283	GACCATGGATGATGCGCGCTGGGGGTATTTCTTTTACAAATGCTGGGATGCTTTTGGGT	2342
Db	2265	GACCATGGATGATGCGCGCTGGGGGTATTTCTTTTACAAATGCTGGGATGCTTTTGGGT	2324
QY	2343	TTTGGAATATGTTACTATATTTCCATATTTTGAACATAAATCGTGTTCCTATATAAATTTTC	2402
Db	2325	TTTGGAATATGTTACTATTTCCATATTTTGAACATAAATCGTGTTCCTATATAAATTTTC	2384
QY	2403	TAGTTTAAGGTTTAAATGCTAGAGTAGCTTAAAAGTTGTCATAGTTTCAGTAGTCAATAG	2462
Db	2385	TAGTTTAAGGTTTAAATGCTAGAGTAGCTTAAAAGTTGTCATAGTTTCAGTAGTCAATAG	2444
QY	2463	GGAGATGAGTTTGTATGCTGTACTATGACGCGTTTAAAGTTAGTGGGTTTGTGATTTTTT	2522
Db	2445	GGAGATGAGTTTGTATGCTGTACTATGACGCGTTTAAAGTTAGTGGGTTTGTGATTTTTT	2504
QY	2523	GTATTGAATATTCGCTGCTGTTTACAAAGTCAGTTTAAAGTACGTTTAAATATTTTAAGTTA	2582
Db	2505	GTATTGAATATTCGCTGCTGTTTACAAAGTCAGTTTAAAGTACGTTTAAATATTTTAAGTTA	2564
QY	2583	TTCTATCTTGGAGATAAATCTGTATGTGCATATCACCGTATTAACAGTTTATTAATGTA	2642
Db	2565	TTCTATCTTGGAGATAAATCTGTATGTGCATATTAACCGGTTATTAACAGTTTATTAATGTA	2624
QY	2643	AACAGAGATTTGGCATGACATGTTCTGTATGTTTTCAGGGAAAAATGCTTTAATGCTTTT	2702
Db	2625	AACAGAGATTTGGCATGACATGTTCTGTATGTTTTCAGGGAAAAATGCTTTAATGCTTTT	2684
QY	2703	TTTCAAGAACTAACACAGTTTATTCCTATATCTAGGATTTTATAGTTTCTCTGAAAGAACTGCTGGTG	2762
Db	2685	TTTCAAGAACTAACACAGTTTATTCCTATATCTAGGATTTTATAGTTTCTCTGAAAGAACTGCTGGTG	2744
QY	2763	TTTGAAGAATAAGAAATGCGATGAAGCGCTAAAATACCAAGAAAGCTTATATCTGAATTTAAG	2822
Db	2745	TTTGAAGAATAAGAAATGCGATGAAGCGCTAAAATACCAAGAAAGCTTATATCTGAATTTAAG	2804

QY	1081	CTCAAAGACCTATTTCATTAACAATASGCTGGGTGGTGGTTTATFAGCCATTTCCATCA	1140
DB	1063	CTCCAAGACCTATTTCATTAACAATASGCTGGGTGGTGGTTTATFAGCCATTTCCATCA	1122
QY	1141	TCAGTTTCCTGCTCTGCTGGGGTTATCTTAGTGCCCTCTCATGAATCGGGTGTTTTTCA	1200
DB	1123	TCAGTTTCCTGCTCTGCTGGGGTTATCTTAGTGCCCTCTCAATGAATCGGGTGTTTTTCA	1182
QY	1201	AATTTCTCTGAGTTTCCTTTGTGSCACTGGCCGTTGGGACTTTGAGTGGTGATGCTTTTT	1260
DB	1183	AATTTCTCTGAGTTTCCTTTGTGSCACTGGCCGTTGGGACTTTGAGTGGTGATGCTTTTT	1242
QY	1261	TACACCTTTCTCCACATTTCTCATGCAAGTCAACCACCTAGTCTATGCCATGAAGAACCAG	1320
DB	1243	TACACCTTTCTCCACATTTCTCATGCAAGTCAACCACCTAGTCTATGCCATGAAGAACCAG	1302
QY	1321	CAATGGAAATGAAGAGAGGCCACTTTTCAGTCACTCTCTTCTCTCAAAACATAGAAGAAA	1380
DB	1303	CAATGGAAATGAAGAGAGGCCACTTTTCAGTCACTCTCTTCTCTCAAAACATAGAAGAAA	1362
QY	1381	GTGCTATTTGATTTCCACGTGGAGGGTCTAACAGCTCTAGGAGCCCTGTATTTCAATGT	1440
DB	1363	GTGCTATTTGATTTCCACGTGGAGGGTCTAACAGCTCTAGGAGCCCTGTATTTCAATGT	1422
QY	1441	TTCTTTGTGAACATGCTCTCCATGATCAAAACAATTTAAAGATAGAAGAAAAAAGAAATC	1500
DB	1423	TTCTTTGTGAACATGCTCTCCATGATCAAAACAATTTAAAGATAGAAGAAAAAAGAAATC	1482
QY	1501	AGAAGAACCTGAATAATGATGATGATGTGGAGATTAAAGACGAGTTGTCCAGATATGAAT	1560
DB	1483	AGAAGAACCTGAATAATGATGATGATGTGGAGATTAAAGACGAGTTGTCCAGATATGAAT	1542
QY	1561	CTCAACCTTTCAACAAATGAGAGAGAAATAGATACAGATGATGAACTGAAGGCTATTTAC	1620
DB	1543	CTCAACCTTTCAACAAATGAGAGAGAAATAGATACAGATGATGAACTGAAGGCTATTTAC	1602
QY	1621	GAGCAGACTCACAGAGCCCTCCACTTTGATTTCTACAGAGCCTGCAGTCITGGAGAAG	1680
DB	1603	GAGCAGACTCACAGAGCCCTCCACTTTGATTTCTACAGAGCCTGCAGTCITGGAGAAG	1662
QY	1681	RAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAG	1740
DB	1663	RAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAG	1722
QY	1741	GGTCAAGAATAAATGGCCATTCACATTTCCACGATFACACTCGGCCAGTCAGACGATFCA	1800
DB	1723	GGTCAAGAATAAATGGCCATTCACATTTCCACGATFACACTCGGCCAGTCAGACGATFCA	1782
QY	1801	TTCCACCACCATGATGACTACCATCATATTTCTTCATCATCACACACCAAAAACCAACATC	1860
DB	1783	TTCCACCACCATGATGACTACCATCATATTTCTTCATCATCACACACCAAAAACCAACATC	1842
QY	1861	CTCAGTCTCAGCCAGCGCTACTCTCGGAGAGACTGAAAGATGCGCGGCTGCACACTT	1920
DB	1843	CTCAGTCTCAGCCAGCGCTACTCTCGGAGAGACTGAAAGATGCGCGGCTGCACACTT	1902
QY	1921	TGGCCTGGATGGTGATTAATGGGTGATGGCTGCACAAATTTCAAGCGATGSCCTAGCAATG	1980
DB	1903	TGGCCTGGATGGTGATTAATGGGTGATGGCTGCACAAATTTCAAGCGATGSCCTAGCAATG	1962
QY	1981	GTGCTGCTTTTACTGAAGCTTATCAAGTGGTTAAAGTACTTCTGTTGCTGTGTTCTGTC	2040
DB	1963	GTGCTGCTTTTACTGAAGCTTATCAAGTGGTTAAAGTACTTCTGTTGCTGTGTTCTGTC	2022
QY	2041	ATGAGTTGCCCTCATGAATTTAGTGACCTTTGCTGTTCTACTAAAGCTGGCATGACCGTTA	2100
DB	2023	ATGAGTTGCCCTCATGAATTTAGTGACCTTTGCTGTTCTACTAAAGCTGGCATGACCGTTA	2082
QY	2101	ACGAGCTGTCCTTTAATGATGTGTCACCATGCTGGCTATCTTTGGAATGGCAACAG	2160
DB	2083	ACGAGCTGTCCTTTAATGATGTGTCACCATGCTGGCTATCTTTGGAATGGCAACAG	2142
QY	2161	GAATTTTTCATTTGCTAATTATGCTGAAATGTTCCTATGTGGATATTTTGCACTTACTGCTG	2220

Db	2143	GAATTTTCATTTGGTCATTATCTCGAAATGTTTCTTATGTGATATTGGCACCTACTGCTG	2202
Qy	2221	CGTTATTCAATGATGTTGCTCTGGTTCATATGGTACCCTGAAATCGCTGCACAATGATGCTA	2280
Db	2203	CGTTATTCAATGATGTTGCTCTGGTTCATATGGTACCCTGAAATCGCTGCACAATGATGCTA	2262
Qy	2281	GTGACAAATGATGTAGCCGCTGGGGGTATTTCTTTTACAAATGCTGGAGCTCTTTGG	2340
Db	2263	GTGACAAATGATGTAGCCGCTGGGGGTATTTCTTTTACAAATGCTGGAGCTCTTTGG	2322
Qy	2341	GTTTTGGAAATATGATCTATTTCCCATATTTGCAATGCTGTTTCGTATATAAATT	2400
Db	2323	GTTTTGGAAATATGATCTATTTCCCATATTTGCAATGCTGTTTCGTATATAAATT	2382
Qy	2401	TCTAGTTAAGTTTAAATGCTAGATAGCTTAAAAGTTGTCATAGTTTCAGTAGGTCAT	2460
Db	2383	TCTAGTTAAGTTTAAATGCTAGATAGCTTAAAAGTTGTCATAGTTTCAGTAGGTCAT	2442
Qy	2461	AGGAGATAGAGTTTGTATGCTGTACTATGACAGCGTTAAAAGTTAGTGGGTTTGTGATTT	2520
Db	2443	AGGAGATAGAGTTTGTATGCTGTACTATGACAGCGTTAAAAGTTAGTGGGTTTGTGATTT	2502
Qy	2521	TGTATTTGAATATGCTGTCTGTGTACAAAGTCAGTTAAAGTCAGCTTTTAATATTTAGT	2580
Db	2503	TTGTATTTGAATATGCTGTCTGTGTACAAAGTCAGTTAAAGTCAGCTTTTAATATTTAGT	2562
Qy	2581	TATTTCTATCTTTGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAAGTTTATTTATG	2640
Db	2563	TATTTCTATCTTTGAGATAAAATCTGTATGTGCAATTCACCGGTATTACCAGTTTATTTATG	2622
Qy	2641	TAAACAAGAGATTGGCATGACATGTTCTCTATGTTTTCAGGGAAAAATGCTTTTAATGCT	2700
Db	2623	TAAACAAGAGATTGGCATGACATGTTCTCTATGTTTTCAGGGAAAAATGCTTTTAATGCT	2682
Qy	2701	TTTTCAAGAACTAACACAGATTATCTTATCTACTGGATTTTAGTGCTCTGGAAGAACTGCTGG	2760
Db	2683	TTTTCAAGAACTAACACAGATTATCTTATCTACTGGATTTTAGTGCTCTGGAAGAACTGCTGG	2742
Qy	2761	TG 2762	
Db	2743	TG 2744	
RESULT 4			
US-10-177-293-263			
; Sequence 263, Application US/10177293			
; Publication NO. US20030124128A1			
; GENERAL INFORMATION:			
; APPLICANT: Lillie, James			
; APPLICANT: Glatt, Karen			
; APPLICANT: Zhao, Xumei			
; APPLICANT: Gannavarpu, Manjula			
; APPLICANT: Kamatkar, Shubhangi			
; APPLICANT: Mertens, Maureen			
; APPLICANT: Myer, Vic			
; APPLICANT: Wang, Youzhen			
; APPLICANT: Xu, Yongyao			
; APPLICANT: Hoersch, Sebastian			
; APPLICANT: Monahan, John			
; APPLICANT: Meyers, Rachel E.			
; APPLICANT: Bast Jr., Robert C.			
; APPLICANT: Hortobagyi, Gabriel N.			
; APPLICANT: Pusztai, Lajos			
; APPLICANT: Meric, Funda			
; APPLICANT: Sahin, Aysegul			
; APPLICANT: Mills, Gordon B.			
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, AND THERAPY OF BREAST CANCER			
; FILE REFERENCE: MRI-038			
; CURRENT APPLICATION NUMBER: US/10/177,293			
; CURRENT FILING DATE: 2002-06-21			
; PRIOR APPLICATION NUMBER: US 60/299,887			

Query Match	78.5%;	Score 2716;	DB 14;	Length 2744;
Best Local Similarity	99.3%;	Pred. No. 0;	Mismatches 0;	Gaps 1;
Matches 2744;	Conservative	0;	Indels 18;	Gaps 1;

QY	1	CTCGTGGCCGAATTCGCGACGAGACCGCGTGTTGCGCCCTGGTAGAGATTCTCGAAGACA	60
Db	1	CTCGTGGCCGAATTCGCGACGAGACCGCGTGTTGCGCCCTGGTAGAGATTCTCGAAGACA	60
QY	61	CCAGTGGGCCGTGTGGAAACCAACCFCGCGCGGTGGCGGCCTGGGACCAACAGGCC	120
Db	61	CCAGTGGGCCGTGTGGAAACCAACCFCGCGCGGTGGCGGCCTGGGACCAACAGGCC	120
QY	121	CGCGAGCAGAAGCGCAATGCGGAGGAGTAATCTGTAATCTTGATCCTGACCTTTGCC	180
Db	121	CGCGAGCAGAAGCGCAATGCGGAGGAGTAATCTGTAATCTTGATCCTGACCTTTGCC	180
QY	181	TCTCTGTCACAAATCCCTTCATGAACTAAAGCAGCTGCTTCCCCAGACCACTGAGA	240
Db	181	TCTCTGTCACAAATCCCTTCATGAACTAAAGCAGCTGCTTCCCCAGACCACTGAGA	240
QY	241	AAATTAGTCGGAATGGGAATCTGSCAATAATGTTGACTTGCAATTTCCACACGCGAAT	300
Db	241	AAATTAGTCGGAATGGGAATCTGSCAATAATGTTGACTTGCAATTTCCACACGCGAAT	300
QY	301	ATCATCTACACAGCTTTTTTACCGCTATGCGAGAAAATAAATCTTTGCTAGTTGAAGGT	360
Db	301	ATCATCTACACAGCTTTTTTACCGCTATGCGAGAAAATAAATCTTTGCTAGTTGAAGGT	360
QY	361	TCAGAAAAATTACTTCAAATATAGGATAGATAAGATTAAAGAATCCATATACACCATG	420
Db	361	TCAGAAAAATTACTTCAAATATAGGATAGATAAGATTAAAGAATCCATATACACCATG	420
QY	421	ACCACGACATCACTCAGACCAACGAGCATCACTCAGACCATGAGCGTCCTCAGACCATG	480
Db	421	ACCACGACATCACTCAGACCAACGAGCATCACTCAGACCATGAGCGTCCTCAGACCATG	480
QY	481	AGCATCACTCAGACCAACGAGCATCACTCAGACCATGATCATCACTCTCACCAATAATCATG	540
Db	481	AGCATCACTCAGACCAACGAGCATCACTCTGAACAT-----AATCATG	522
QY	541	CTGCTTCTGGTAAATAATGAGCGAAAAGCTTTTGGCCAGACCATCACTCAGATGAGTTCAG	600
Db	523	CTGCTTCTGGTAAATAATGAGCGAAAAGCTTTTGGCCAGACCATCACTCAGATGAGTTCAG	582
QY	601	GTAAGAATCTCTAGAACAGCGAGGGGAAAGAGCTCACCAGCACGAGACATGCCAGTGGTA	660
Db	583	GTAAGAATCTCTAGAACAGCGAGGGGAAAGAGCTCACCAGCACGAGACATGCCAGTGGTA	642
QY	661	GAAGGAATGTCAAGGACAGCTGTAGTGCTAGTGAAGTACCTCAACTGTGTACAACTG	720
Db	643	GAAGGAATGTCAAGGACAGCTGTAGTGCTAGTGAAGTACCTCAACTGTGTACAACTG	702
QY	721	TCTCTGAGGAATCTCACTTTCTAGAGACAAATAGAGACTCCAAGACCTGGAAAACTTCCC	780


```

; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13805
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013,
; LOCATION: 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023,
; LOCATION: 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-13805

Query Match      20.1%; Score 694.2; DB 14; Length 1032;
Best Local Similarity 90.0%; Pred. No. 3.1e-166;
Matches 800; Conservative 0; Mismatches 83; Indels 6; Gaps 5;

QY      351 GTTGAGGGTTTCAGAAAATTACTTCAAATATATAGSCATAGATTAAGATTAAAGAATCCAT 410
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      4   GTCCGAGGGTTAGAAAATTACTT-AAAATATAGCATAGATAGATTAAAGAATCCAT 62

QY      411 ATACACATGACCGACGACCATCACTCAGACCACGAGCATCAC-TCAGACCATGAGC-GTC 468
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      53 ATACACCATGACCGACGACCATCACTCAGACCACGAGCATCACGTCAGACCATGACGGTC 122

QY      469 ACTCAGACCATGAGC-AFTCACTCAGACCACGAGCATCACTCTGACCATGATCATCACTCT 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      123 ACTCAGACCATGAGCAATCACTCAGACCACGAGCATCACTCTGACCATGATCATCACTCT 182

QY      528 CACCATATCATGCTGCTTCTGTTAAATATAGCGAAGAAAGCTCTTTGCCACGACCATGAC 587
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      183 CACCATATCATGCTGCTTCTGTTAAATATAGCGAAGAAAGCTCTTTGCCACGACCATGAC 242

QY      588 TCAGATAGTTCAGGTAAAGATCTCTAGAAACAGCCAGGGGAAAGGAGCTCACCGACGACAA 647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      243 TCAGATAGTTCAGGTAAAGATCTCTAGAAACAGCCAGGGGAAAGGAGCTCACCGACGACAA 302

QY      648 CATGCCAGTGGTAGAAGGAATGTCAAGACAGCTGTAGTGTCTAGAGACAATAGAGACTCCAGCT 707
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      303 CATGCCAGTGGTAGAAGGAATGTCAAGACAGCTGTAGTGTCTAGAGACAATAGAGACTCCAGCT 362

QY      708 GTGTACAAACACTGTCTCTGAAAGNACTCACTCTTCTAGAGACAATAGAGACTCCAGACT 767
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      363 GTGTACAAACACTGTCTCTGAAAGNACTCACTCTTCTAGAGACAATAGAGACTCCAGACT 422

QY      768 GGAATAACTCTTCCCCAAAGATGTAAAGCAGCTCCACTCCACCCAGTGTCACTCAAAAGAGC 827
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      423 GGAATAACTCTTCCCCAAAGATGTAAAGCAGCTCCACTCCACCCAGTGTCACTCAAAAGAGC 482

QY      828 CGGGTGAGCCGGCTGGCTGTGTAGAAACAAATGAATCTGTAGTGTAGCCGCCGAAAAAGGC 887
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      483 CGGGTGAGCCGGCTGGCTGTGTAGAAACAAATGAATCTGTAGTGTAGCCGCCGAAAAAGGC 542

QY      888 TTATGTATTTCCAGAAAACACAATATGAATATCTCTAGGAGTGTTCATATGCATCAAGCTA 947
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      543 TTATGTATTTCCAGAAAACACAATATGAATATCTCTAGGAGTGTTCATATGCATCAAGCTA 602

QY      948 CTGACATCTCATGGCATGGGCATCCAGGTTTCGCGTGAATGCAACAGAGTTTCAACTATCTC 1007
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      603 CTGACATCTCATGGCATGGGCATCCAGGTTTCGCGTGAATGCAACAGAGTTTCAACTATCTC 662

QY      1008 TGTCCGCCCATCATCAACCAATTTGATGCTAGATCTTCTGTGATTCATCAAGTCAAAAG 1067
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      663 TGTCCGCCCATCATCAACCAATTTGATGCTAGATCTTCTGTGATTCATCAAGTCAAAAG 722

QY      1068 AAGCTGTAATCCCTCCAAAGACCTATTCATTAATAATAGCCTGGGTGGTGGTTTATA 1127
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 430 GATCGACTGAGCGCTATTACAGACGAGCTCACAGAGCCCTCCACCTTGATCTCTCAG 489
QY 1659 CAGCCTGCGAGCTTGGAGAGAGAGGTCATGATAGCTCATGCTCATCCACAGGAATC 1718
Db 490 CAGCCTGCGAGCTTGGAGAGAGAGGTCATGATAGCTCATCCACAGGAATC 549
QY 1719 TACAATGAATATGTACCCAGAGGTCGAAGATTAATGCCATTCACATTTCCACA 1774
Db 550 TACAATGAATATGTACCCAGAGGTCGAAGATTAATGCCATTCACATTTCCACA 605

RESULT 9
US-10-198-846-11215
; Sequence 11215, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR FILING DATE: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11215
; LENGTH: 1598
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1594, 1595, 1596, 1597, 1598
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-11215

Query Match 16.7%; Score 578.4; DB 14; Length 1598;
Best Local Similarity 91.3%; Pred. No. 1.5e-136;
Matches 725; Conservative 0; Mismatches 56; Indels 13; Gaps 10;

QY 2245 TTGATATGCTACTGGAATGCTGCACAAATGCTAGTACGATGGATGGATGGATGGG 2304
Db 682 TTTTAAAGGTACCTGGAATGCTGCACAAATGCTAGTACGATGGATGGATGGG 741
QY 2305 GGTATTTCCTTTTACAGAAATGCTGGATGCTTTTGGGTTTGGAAATGCTTACTTATT 2364
Db 742 GGTATTTCCTTTTACAGAAATGCTGGATGCTTTTGGGTTTGGAAATGCTTACTTATT 799
QY 2365 CCATATTGAACATAAAATGCTGTTTCGTATATAAATTCAGTAAAGTTTAAATGCTAGA 2424
Db 800 CCATATTGAACATAAAATGCTGTTTCGTATATAAATTCAGTAAAGTTTAAATGCTAGA 859
QY 2425 GTAGCTTAAAGTTTGCATAGTTTCAGTAGTGCATAGGGATGAGTTTGTATGCTGTA 2484
Db 860 GTAGCTTAAAGTTTGCATAGTTTCAGTAGTGCATAGGGATGAG-TTGTATGCTGTA 918
QY 2485 CTATGACGCGTTTAAAGTTTGTAGTTTGTGATTTTGTATTAATGCTGCTGTT 2544
Db 919 CTATGACGCGTTTAAAGTTTGTAGTTTGTGATTTTGTATTAATGCTGCTGTT 976
QY 2545 ACAAGTCAGTTTAAAGTTTGTAGTTTGTATTAATGCTGCTGTTTGTAGGAGATAAATCT 2604
Db 977 ACAAGTCAGTTTAAAGTTTGTAGTTTGTATTAATGCTGCTGTTTGTAGGAGATAAATCT 1036
QY 2605 GTATGTCATATCCCGGTATACAGTTTATATGTAACAAGAGATTTGGCATGACAT 2664
Db 1037 GTATGTCATATCCCGGTATACAGTTTATATGTAACAAGAGATTTGG-ATGACAT 1095

QY 2665 GTTCTGTATGTTTTCAGGAAAAATGCTTTAATGCTTTTTCAAGAACTAACACAGTTATT 2724
Db 1096 GTTCTGTATGTTTTCAGGAAAAATGCTTTAATGCTTTTTCAAGAACTAACACAGTTATT 1155
QY 2725 CTTATACCTGGATTTTAGTCTCTGAAGAACTGCTGCTGTTTAGGAATAAGAATGTGCATG 2784
Db 1156 CTTATACCTGGATTTTAGTCTCTGAAGAACTG-TGGTGTTTAGGAATAAGAATGTGCATG 1214
QY 2785 AGCCTTAAATACCAAGAAAGCTTATACCTGAATTTAAGCAAGAAATAAGAGAGAAAGA 2844
Db 1215 AGCCTTAAATACCAAGAAAGCTTATACCTGAATTTAAGCAAGAAATAAGAGAGAAAGA 1274
QY 2845 GAAGATCTGGAATTTGGGAGGATAGATTTCTTATAAAATCACAAAATTTGTTGTAAA 2904
Db 1275 GAAGATCTGAGAA-TGGGGAGGATAGA-TCTTTATAAAATCACAAAATTCGT--GTAA 1330
QY 2905 TTAGAGGGGAGAAATTTAGAAATTAAGTA-TAAAAAGGAGAAATTTAGTAGAGTACATC 2963
Db 1331 TTAGAGGGGAGAAATTCAGAACTTACGTAATACAGGGCGGAATTTAGTCTTTGAAGTCCATC 1390
QY 2964 ATTAACATTTTGTCCAGGATTTTCCCGTAAACAGCTAGTACACACTCTCATATACTA 3023
Db 1391 CTCACACTTCTGTGC-GGATTTTCCGTTAAAAACGTTGTGGCCCTTTCTATCTAATT 1449
QY 3024 ATTAGTGTACATTT 3037
Db 1450 TCGTGTCTTACTT 1463

RESULT 10
US-09-867-701-2021
; Sequence 2021, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2021
; LENGTH: 483
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(483)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-2021

Query Match 11.7%; Score 404.8; DB 10; Length 483;
Best Local Similarity 92.1%; Pred. No. 1.2e-92;
Matches 444; Conservative 0; Mismatches 35; Indels 3; Gaps 2;

QY 2541 TGTTCACAAAGTCAGTTAAAGTACGTTTAAATTTAAATTTAACTTATCTATCTTGGAGATAA 2600
Db 2 TGTTCACAAAGTCAGTTAAAGTACGTTTAAATTTAAATTTAACTTATCTATCTTGGAGATAA 61
QY 2601 ATCTGTATGTCGAATTCACCGGATTAACAGTTTATATGTAACAAGAGATTTGGCATG 2660
Db 62 ATCTGTATGTCGAATTCACCGGATTAACAGTTTATATGTAACAAGAGATTTGGCATG 121
QY 2661 ACATGTTCTGTATGTTTTCAGGAAAAATGCTTTAATGCTTTTTCAAGAACTAACACAGT 2720
Db 122 ACATGTTCTGTATGTTTTCAGGAAAAATGCTTTAATGCTTTTTCAAGAACTAACACAGT 181
QY 2721 TATTCCTATACCTGGAATTTTAGTCTCTGAAGAACTGCTGGTGTTTTAGAATAAGAATGTG 2780
Db 182 TATTCCTATACCTGGAATTTTAGTCTCTGAAGAACTGCTGGTGTTTTAGAATAAGAATGTG 241


```
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-827-669-78

Query Match      10.4%; Score 360; DB 9; Length 385;
Best Local Similarity 98.2%; Pred. No. 2.7e-81;
Matches 375; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 2453 TAGTCATAGGAGATGAGTTGCTGCTACTATGACAGCGTTTAAAGTTAGTGGGTTT 2512
Db 1 TCGGTCATAGGAGAG--GATTGCTATGCTACTATGACAGCGTTTAAAGTTAGTGGGTTT 58

QY 2513 TGTGATTTTGTATTCGAATATTTGCTGCTGCTTACAAAGTCAGTTAAAGGTACGTTTAAAT 2572
Db 59 TGTGATTTTGTATTCGAATATTTGCTGCTGCTTACAAAGTCAGTTAAAGGTACGTTTAAAT 118

QY 2573 ATTTAAGTTATTCATCTCTTTGGAGATAAAATCTGTATGTGCAATTCACOGGTATTACCACT 2632
Db 119 ATTTAAGTTATTCATCTCTTTGGAGATAAAATCTGTATGTGCAATTCACOGGTATTACCACT 178

QY 2633 TTATATGCTAAACAAGAGATTGGCATGCACATGCTCTGTATCTTTACAGGAAAAATGCTCT 2692
Db 179 TTATATGCTAAACAAGAGATTGGCATGCACATGCTCTGTATCTTTACAGGAAAAATGCTCT 238

QY 2693 TTAATGCTTTTTCAGAACTAACACAGTTATTCCTACTACTGATTTTAGGTCCTCTGAAGA 2752
Db 239 TTAATGCTTTTTCAGAACTAACACAGTTATTCCTACTACTGATTTTAGGTCCTCTGAAGA 298

QY 2753 ACTGCTGGTGTAGGATATAGAAATGTGCATGAGCCTAAATACCAAGAAAGCTTATAC 2812
Db 299 ACTGCTGGTGTAGGATATAGAAATGTGCATGAGCCTAAATACCAAGAAAGCTTATAC 358

QY 2813 TGAATTTAAGCAAGAAATATA 2834
Db 359 TGAATTTAAGCAAGAAATATA 380
```

Search completed: September 12, 2003, 22:12:41
Job time : 804.001 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 16:38:39 ; Search time 126.286 Seconds
(without alignments)
7926.909 Million cell updates/sec

Title: US-09-642-034-4
Perfect score: 2268
Sequence: 1 atggcgagggaagtatctgt.....tggttgatatataattctgt 2268

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1263	55.7	1310	1 US-08-311-023-1	Sequence 1, Appli
2	1263	55.7	2404	1 US-08-311-023-3	Sequence 3, Appli
3	110.6	4.9	1888	4 US-09-599-360B-38	Sequence 38, Appli
c 4	59.8	2.6	285	3 US-08-990-571-64	Sequence 64, Appli
c 5	59.8	2.6	285	4 US-09-528-784A-64	Sequence 64, Appli
c 6	59.8	2.6	285	4 US-09-569-098A-64	Sequence 64, Appli
c 7	59.8	2.6	342	3 US-08-990-571-65	Sequence 65, Appli
c 8	59.8	2.6	342	4 US-09-528-784A-65	Sequence 65, Appli
c 9	59.8	2.6	342	4 US-09-569-098A-65	Sequence 65, Appli
c 10	59.8	2.6	351	3 US-08-990-571-60	Sequence 60, Appli
c 11	59.8	2.6	351	4 US-09-528-784A-60	Sequence 60, Appli
c 12	59.8	2.6	351	4 US-09-569-098A-60	Sequence 60, Appli
c 13	59.8	2.6	356	3 US-08-990-571-63	Sequence 63, Appli
c 14	59.8	2.6	356	4 US-09-528-784A-63	Sequence 63, Appli
c 15	59.8	2.6	356	4 US-09-569-098A-63	Sequence 63, Appli
c 16	59.8	2.6	358	3 US-08-990-571-58	Sequence 58, Appli
c 17	59.8	2.6	358	4 US-09-528-784A-58	Sequence 58, Appli
c 18	59.8	2.6	358	4 US-09-569-098A-58	Sequence 58, Appli
c 19	59.8	2.6	363	3 US-08-990-571-66	Sequence 66, Appli
c 20	59.8	2.6	363	3 US-08-990-571-67	Sequence 67, Appli
c 21	59.8	2.6	363	4 US-09-528-784A-66	Sequence 66, Appli
c 22	59.8	2.6	363	4 US-09-528-784A-67	Sequence 67, Appli
c 23	59.8	2.6	363	4 US-09-569-098A-66	Sequence 66, Appli
c 24	59.8	2.6	363	4 US-09-569-098A-67	Sequence 67, Appli
c 25	59.8	2.6	409	3 US-08-990-571-59	Sequence 59, Appli
c 26	59.8	2.6	409	4 US-09-528-784A-59	Sequence 59, Appli
c 27	59.8	2.6	409	4 US-09-569-098A-59	Sequence 59, Appli

c 28	59.8	2.6	410	3 US-08-990-571-61	Sequence 61, Appli
c 29	59.8	2.6	410	4 US-09-528-784A-61	Sequence 61, Appli
c 30	59.8	2.6	410	4 US-09-569-098A-61	Sequence 61, Appli
c 31	59.8	2.6	416	3 US-08-990-571-62	Sequence 62, Appli
c 32	59.8	2.6	416	4 US-09-528-784A-62	Sequence 62, Appli
c 33	59.8	2.6	416	4 US-09-569-098A-62	Sequence 62, Appli
c 34	59.8	2.6	1271	3 US-08-845-258-5	Sequence 5, Appli
c 35	59.8	2.6	1271	3 US-08-845-258-42	Sequence 5, Appli
c 36	59.8	2.6	1271	3 US-08-990-571-5	Sequence 5, Appli
c 37	59.8	2.6	1271	3 US-08-990-571-42	Sequence 5, Appli
c 38	59.8	2.6	1271	4 US-08-723-142A-5	Sequence 5, Appli
c 39	59.8	2.6	1271	4 US-08-723-142A-42	Sequence 5, Appli
c 40	59.8	2.6	1271	4 US-09-528-784A-5	Sequence 5, Appli
c 41	59.8	2.6	1271	4 US-09-528-784A-42	Sequence 5, Appli
c 42	59.8	2.6	1271	4 US-09-569-098A-5	Sequence 5, Appli
c 43	59.8	2.6	1271	4 US-09-569-098A-42	Sequence 5, Appli
c 44	59.8	2.6	1821	3 US-08-845-258-6	Sequence 6, Appli
c 45	59.8	2.6	1821	3 US-08-990-571-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-311-023-1
; Sequence 1, Application US/083111023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; TITLE OF INVENTION: BREAST TUMOURS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1310 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1293
US-08-311-023-1

Query Match 55.7%; Score 1263; DB 1; Length 1310;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1285; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY	1102	ACTTTGAGTGTGARGCTTTTTTACACCCTTCCTCCACATCTTCAICGAAGHCACACCAT	1167
DB	121	ACTTTGAGTGTGARGCTTTTTTACACCCTTCCTCCACATCTTCAICGAAGTCACACCAT	180
QY	1162	ASTCATAGCCATGABAGAACGACGAANTGGAAATGAAAGACAGGCCACTTTTCAGTCANCTG	1221
DB	181	AGTCATAGCCATGABAGAACGACGAANTGGAAATGAAAGACAGGCCACTTTTCAGTCANCTG	240
QY	1222	TCTTCTCAAACAATAGAGAAGTGCGCTATTTTGGATTCTCACGTGGAGGGSTCTAACAGCT	1281
DB	241	TCITCTCAAACAATAGAGAAGTGCGCTATTTTGGATTCTCACGTGGAGGGSTCTAACAGCT	300
QY	1282	CTAGGAGGCCTGTATTTTCATGTTCTTGTTGAACTATGCTCCTCACATTTGATCAACAATTT	1341
DB	301	CTAGGAGGCCTGTATTTTCATGTTCTTGTTGAACTATGCTCCTCACATTTGATCAACAATTT	360
QY	1342	AAAGATAAGAAGAAAAGAAATCAAGAAGAACTGAAAAATGATGATGTGAGAGATTAAAG	1401
DB	361	AAAGATAAGAAGAAAAGAAATCAAGAAGAACTGAAAAATGATGATGTGAGAGATTAAAG	420
QY	1402	AAGCAGTTCTCCAAGTAATGAAATCTCAACTTTCAACAATGAGGAGAAAGTAGATACAGAT	1461
DB	421	AAGCAGTTCTCCAAGTAATGAAATCTCAACTTTCAACAATGAGGAGAAAGTAGATACAGAT	480
QY	1462	GATCGAAGTGAAGGCTATTTACGAGCAGACTCACAAGAGCCCTCCACATTTGATTCTCTCAG	1521
DB	481	GATCGAAGTGAAGGCTATTTACGAGCAGACTCACAAGAGCCCTCCACATTTGATTCTCTCAG	540
QY	1522	CAGCTGCGAGCTTGTGGAAGAAAGAGGTCATGATGAGCTCATGCTCATCCACAGGAATC	1581
DB	541	CAGCTGCGAGCTTGTGGAAGAAAGAGGTCATGATGAGCTCATGCTCATCCACAGGAATC	600
QY	1582	TACAATGATATGTACCCAGAGGGTGCAGAATAAATGCCATTACATTTCCACGATACA	1641
DB	601	TACAATGATATGTACCCAGAGGGTGCAGAATAAATGCCATTACATTTCCACGATACA	660
QY	1642	CTGGCCAGTCAGACGATCTCATTCCACCACCATGATGACTACCATCATATTTCTCCCAT	1701
DB	661	CTGGCCAGTCAGACGATCTCATTCCACCACCATGATGACTACCATCATATTTCTCCCAT	720
QY	1702	CACCACCACCAAAACCACTCTCCAGCTCAGCCAGCCCTACTCTGGGAGAGCTG	1761
DB	721	CACCACCACCAAAACCACTCTCCAGCTCAGCCAGCCCTACTCTGGGAGAGCTG	780
QY	1762	AAAGATCGCGGCTCGCCACTTTGGCCCTGGATGGTGATAATGGGTGATGGCCCTGCACAA	1821
DB	781	AAAGATCGCGGCTCGCCACTTTGGCCCTGGATGGTGATAATGGGTGATGGCCCTGCACAA	840
QY	1822	TTACGCGATGSCCTAGCAATTTGGTGTGCTTTTACTGAAAGCTTATFCAGTGGTTAAGT	1881
DB	841	TTACGCGATGSCCTAGCAATTTGGTGTGCTTTTACTGAAAGCTTATFCAGTGGTTAAGT	900
QY	1882	ACTTCTGTCTGTGTTCTGTCAATGAGTTCCTCATGAAATAGTGTGACTTTGCTGTCTTA	1941
DB	901	ACTTCTGTCTGTGTTCTGTCAATGAGTTCCTCATGAAATAGTGTGACTTTGCTGTCTTA	960
QY	1942	CTAAGGCTGGCATGACCGTTAAGCAGGCTGTCCCTTATATGCAATTTGTCAGCCATGCTG	2001
DB	961	CTAAGGCTGGCATGACCGTTAAGCAGGCTGTCCCTTATATGCAATTTGTCAGCCATGCTG	1020
QY	2002	CGGTATCTTGGAAATGGCAACAGGAATTTTCATTGGTTCATATGCTGAAAATGTTCTATG	2061
DB	1021	CGGTATCTTGGAAATGGCAACAGGAATTTTCATTGGTTCATATGCTGAAAATGTTCTATG	1080
QY	2062	TGGATATTTGCCATWTACTGTGGCTTATTCATGATATGTTGCTCTGGTGTGATFGWACT	2121
DB	1081	TGGATATTTGCCATWTACTGTGGCTTATTCATGATATGTTGCTCTGGTGTGATFGWACT	1140
QY	2122	GAAATGCTGCACATATGCTACTGCACATGGATGACCGCTGTGGGTTATTTCTTTTA	2181
DB	1141	GAAATGCTGCACATATGCTACTGCACATGGATGACCGCTGTGGGTTATTTCTTTTA	1200
QY	2182	CAGAAATGCTGGGATGCTTTTGGGTTTTTGGAAATATGTAATCTTATTTCCATATTGAAAT	2241

Db		1201 CAGAATGCTGGATGCTTTTGGGTATTGAATAATTACTTA-TTCCATATTTGAACAT	1259
QY		2242 AAAATCGTGTTTTGGTATAAAATTTCTAG	2268
Db		1260 AAAATCGTG-TTCGTATAAAATTTCTAG	1285
RESULT 3			
US-09-599-360B-38			
; Sequence 38, Application US/09599360B			
; Patent No. 6548633			
; GENERAL INFORMATION:			
; APPLICANT: Dumas Milne Edwards, J.B.			
; APPLICANT: Bouquelieret, L.			
; APPLICANT: Jobert, S.			
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides			
; FILE REFERENCE: GENSET.050CP3			
; CURRENT APPLICATION NUMBER: US/09/599,360B			
; CURRENT FILING DATE: 2000-06-21			
; PRIOR APPLICATION NUMBER: 60/113,686			
; PRIOR FILING DATE: 1998-12-22			
; PRIOR APPLICATION NUMBER: 60/141,032			
; PRIOR FILING DATE: 1999-06-25			
; PRIOR APPLICATION NUMBER: 09/469,099			
; PRIOR FILING DATE: 1999-12-21			
; NUMBER OF SEQ ID NOS: 123			
; SOFTWARE: Patent.pm			
; SEQ ID NO 38			
; LENGTH: 1888			
; TYPE: DNA			
; ORGANISM: Homo Sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 139..1389			
; NAME/KEY: sig_peptide			
; LOCATION: 139..198			
; OTHER INFORMATION: Von Heijne matrix			
; OTHER INFORMATION: score 5.00			
; OTHER INFORMATION: seq HLLAGFCVVVLG/VV			
; NAME/KEY: polyA_signal			
; LOCATION: 1854..1859			
; NAME/KEY: polyA_site			
; LOCATION: 1873..1888			
US-09-599-360B-38			
Query Match 4.9%; Score 110.6; DB 4; Length 1888;			
Best Local Similarity 58.4%; Pred.No. 3, 9e-21;			
Matches 211; Conservative 0; Mismatches 149; Indels 1; Gaps 1;			
QY	1771	GGCCTGCCACTTTGGCCTGGATGGTGATPAATGGGTGATGCCCTGCACAATTTACGGAT	1830
Db	1273	GGTGGCACTGATATCACGTGGATGGTGCTCTCCCTGGGAGATGGCTACACAACCTCATGAT	1332
QY	1831	GGCTAGCAATTTGGTGTCTTTTACTGAAGCCTTATCACTGTGTTTAAGTACTTCTGTT	1890
Db	1333	GGCTGGCCAATAGTGCTGCTTCTCTGATGGCTTCTC-CGGGCCCTCAGTACCACCTTA	1391
QY	1891	GTGTGTTCCTGTCATGAGTTGCCCTCATGAATTAGTGACTTTGCTGTCTACTAAAGCT	1950
Db	1392	CGGGTCTTCTGCATAGCTGCCACGAACTGGGTGACTTTGGCATGCTGCTCCAGTCA	1451
QY	1951	GCATCCCGTTAGCAGGCGTCCCTTATATGATGATGTCAGCCTGCTGGCGTATCTT	2010
Db	1452	GGGCTGTCTCTTTCGGGGCTGCTGCTGCTGAGCCTCGTGTCTGGAGCCCTGGGATTTGGGG	1511
QY	2011	GGAAATGGCAACAGGAATTTTCATGTGCTATTATGCTGAAAATGTTTCTATGTGGATATT	2070
Db	1512	GTGCAAGCTCTGGGGTGGGGCTACGCTGGGCCCTGTCCCCTCACTCCCTGGGTGTTT	1571
QY	2071	GCATTACTGCTGGCTTATCATGATGTTGCTCTGGTTGATATGFGPACTGAATGCTG	2130
Db	1572	GGGGTCACTGCTGGGGTCTTCCCTGATGTGGCCCTTGTGGACATGCTACACGCTGCTT	1631

OY 2131 C 2131
Db 1632 C 1632

RESULT 4

US-08-990-571-64/c
; Sequence 64, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 285 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-990-571-64

Query Match 2.6%; Score 59.8; DB 3; Length 285;
Best Local Similarity 65.2%; Pred. No. 3.2e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
OY 268 ATCCATATACACCATGAGCATCTCAGACCATCATCTCAGACCATCATCTCAGACCATGAG 327
Db 192 AACCAAGTTCCACTAGGCCACCATGCTTCACTAGGCCACCATGCTTCACTAGGCCACCATG 133
OY 328 CGTCACTAGACCATGAGCATCTCAGACCATGAGCATCTCAGACCATCATCTCAGACCATGAG 387
Db 132 CTTCAGTGGCCACCATGCTTCACTAGGCCACCATGCTTCACTAGGCCACCATGCTTCACT 73
OY 388 TCTACCATATCAT 402
Db 72 TAGGCCACCATGCTT 58

RESULT 5

US-08-528-784A-64/c
; Sequence 64, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-528-784A-64

Query Match 2.6%; Score 59.8; DB 4; Length 285;
Best Local Similarity 65.2%; Pred. No. 3.2e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
OY 268 ATCCATATACACCATGAGCATCTCAGACCATCATCTCAGACCATGAGCATCTCAGACCATGAG 327
Db 192 AACCAAGTTCCACTAGGCCACCATGCTTCACTAGGCCACCATGCTTCACTAGGCCACCATG 133
OY 328 CGTCACTAGACCATGAGCATCTCAGACCATGAGCATCTCAGACCATCATCTCTGACCATGATCATCAC 387
Db 132 CTTCAGTGGCCACCATGCTTCACTAGGCCACCATGCTTCACTAGGCCACCATGCTTCACT 73
OY 388 TCTACCATATCAT 402
Db 72 TAGGCCACCATGCTT 58

RESULT 6

US-09-569-098A-64/c
; Sequence 64, Application US/09569098A
; Patent No. 6569433
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C5
; CURRENT APPLICATION NUMBER: US/09/569,098A
; CURRENT FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-569-098A-64

Query Match 2.6%; Score 59.8; DB 4; Length 285;
Best Local Similarity 65.2%; Pred. No. 3.2e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
OY 268 ATCCATATACACCATGAGCATCTCAGACCATCATCTCAGACCATGAGCATCTCAGACCATGAG 327
Db 192 AACCAAGTTCCACTAGGCCACCATGCTTCACTAGGCCACCATGCTTCACTAGGCCACCATG 133
OY 328 CGTCACTAGACCATGAGCATCTCAGACCATGAGCATCTCAGACCATGAGCATCTGACCATGATCATCAC 387
Db 132 CTTCAGTGGCCACCATGCTTCACTAGGCCACCATGCTTCACTAGGCCACCATGCTTCACT 73
OY 388 TCTACCATATCAT 402
Db 72 TAGGCCACCATGCTT 58

RESULT 7

US-08-990-571-65/c


```
Query Match      2.6%; Score 59.8; DB 4; Length 351;
Best Local Similarity 65.2%; Pred. No. 3.6e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

```

1  RESULT 13
2  US-89-571-63/c
3  ; Sequence 63, Application US/08990571
4  ; Patent No. 6214971
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Reed, Steven G. et al
9  ;
10 ; TITLE OF INVENTION: COMPOUNDS AND
11 ;
12 ; NUMBER OF SEQUENCES: 79
13 ;
14 ; CORRESPONDENCE ADDRESS:
15 ;
16 ; ADDRESS: SEED AND BERRY
17 ;
18 ; STREET: 6300 Columbia Center,
19 ;
20 ; CITY: Seattle
21 ;
22 ; STATE: Washington
23 ;
24 ; COUNTRY: USA
25 ;
26 ; ZIP: 98104
27 ;
28 ; COMPUTER READABLE FORM:
29 ;
30 ; MEDIUM TYPE: Floppy disk
31 ;
32 ; COMPUTER: IBM PC compatible
33 ;

```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/990,571
FILING DATE: 11-DEC-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 356 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-990-571-63

Query Match 2.6%; Score 59.8; DB 3; Length 356;
Best Local Similarity 65.2%; Pred. No. 3.7e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 268 ATCCATATACACCATGACGACCATCTCAGACCATCTCAGACCATGAG 327
DB 210 AACCAGTTCACCTAGGCCACCATCTCAGTACCTAGGCCACCATCTCAGGCCACCATG 151
QY 328 CGTCACTCAGACCATGACGACCATCTCAGACCATCTCAGACCATGATCATCAC 387
DB 150 CTTCACTAGGCCACCATCTCAGTACCTAGGCCACCATCTCAGGCCACCATCTTAC 91
QY 388 TCTCACCATAATCAT 402
DB 90 TAGGCCACCATGCTT 76

RESULT 14
US-09-528-784A-63/C
Sequence 63, Application US/09528784A
Patent No. 6451315
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Sleath, Paul R.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
FILE REFERENCE: 210121.426C4
CURRENT APPLICATION NUMBER: US/09/528,784A
CURRENT FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 90
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 356
TYPE: DNA
ORGANISM: Babesia microti
US-09-528-784A-63

Query Match 2.6%; Score 59.8; DB 4; Length 356;
Best Local Similarity 65.2%; Pred. No. 3.7e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 268 ATCCATATACACCATGACGACCATCTCAGACCATCTCAGACCATGAG 327
DB 210 AACCAGTTCACCTAGGCCACCATCTCAGTACCTAGGCCACCATCTCAGGCCACCATG 151
QY 328 CGTCACTCAGACCATGACGACCATCTCAGACCATCTCAGACCATGATCATCAC 387
DB 150 CTTCACTAGGCCACCATCTCAGTACCTAGGCCACCATCTCAGGCCACCATCTTAC 91

QY 388 TCTCACCATAATCAT 402
DB 90 TAGGCCACCATGCTT 76
RESULT 15
US-09-569-098A-63/C
Sequence 63, Application US/09569098A
Patent No. 6569433
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Sleath, Paul R.
APPLICANT: McNeill, Patricia D.
APPLICANT: Homer, Mary
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
FILE REFERENCE: 210121.426C5
CURRENT APPLICATION NUMBER: US/09/569,098A
CURRENT FILING DATE: 2000-05-10
NUMBER OF SEQ ID NOS: 110
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 356
TYPE: DNA
ORGANISM: Babesia microti
US-09-569-098A-63

Query Match 2.6%; Score 59.8; DB 4; Length 356;
Best Local Similarity 65.2%; Pred. No. 3.7e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 268 ATCCATATACACCATGACGACCATCTCAGACCATCTCAGACCATGAG 327
DB 210 AACCAGTTCACCTAGGCCACCATCTCAGTACCTAGGCCACCATCTCAGGCCACCATG 151
QY 328 CGTCACTCAGACCATGACGACCATCTCAGACCATCTCAGACCATGATCATCAC 387
DB 150 CTTCACTAGGCCACCATCTCAGTACCTAGGCCACCATCTCAGGCCACCATCTTAC 91
QY 388 TCTCACCATAATCAT 402
DB 90 TAGGCCACCATGCTT 76

Search completed: September 12, 2003, 16:55:42
Job time : 128.286 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 16:50:35 ; Search time 518.999 Seconds
 (without alignments)
 10609.165 Million cell updates/sec

Title: US-09-642-034-4
 Perfect score: 2268
 Sequence: 1 atggcaggaagtattctgt.....tggttgataaattctag 2268

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 1632420 seqs, 1213878141 residues

Total number of hits satisfying chosen parameters: 3264840

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2222	98.0	2744	14	US-10-171-581-170
2	2222	98.0	2744	14	US-10-171-293-263
3	2220.4	97.9	3537	14	US-10-198-846-10007
4	2198	96.9	3586	14	US-10-176-847-51
5	1141.4	50.3	1193	10	US-09-925-300-591
6	900.4	39.7	1419	14	US-10-198-846-10321
7	694.2	30.6	1032	14	US-10-198-846-13805
8	578.4	25.5	606	14	US-10-198-846-8490
9	335.6	14.8	366	11	US-09-918-995-30283
10	331.8	14.6	905	14	US-10-198-846-7874
11	319.2	14.1	337	10	US-09-920-455-250
12	319	14.1	4417	12	US-09-814-353-22020
13	319	14.1	5330	9	US-09-789-561-14
c 14	268	11.8	721	14	US-10-198-846-9288
c 15	253	11.2	504	14	US-10-198-846-1455
16	241.2	10.6	625	14	US-10-198-846-10373

c 17	240.8	10.6	469	14	US-10-198-846-408	Sequence 408, App
18	235	10.4	635	14	US-10-198-846-7649	Sequence 7649, App
19	146.8	6.5	479	14	US-10-198-846-1957	Sequence 1957, App
20	139	6.1	1598	14	US-10-198-846-11215	Sequence 11215, A
c 21	132	5.8	550	14	US-10-198-846-5880	Sequence 5880, App
22	128.4	5.7	751	14	US-10-198-846-4814	Sequence 4814, App
23	127.6	5.6	4573	10	US-09-880-107-1649	Sequence 1649, App
24	127.6	5.6	4573	12	US-09-960-706-489	Sequence 489, App
25	126	5.6	4698	12	US-10-240-965-193	Sequence 193, App
26	118.2	5.2	472	14	US-10-198-846-235	Sequence 235, App
27	99.2	4.4	480	12	US-09-814-353-12715	Sequence 12715, A
28	99.2	4.4	481	12	US-09-814-353-1751	Sequence 1751, App
29	96.6	4.3	478	11	US-09-918-995-6431	Sequence 6431, App
30	96.4	4.3	582	10	US-09-796-692-8536	Sequence 8536, App
31	96.4	4.3	582	14	US-10-040-862-8536	Sequence 8536, App
32	88.4	3.9	1848	11	US-09-945-527-29	Sequence 29, Appl
c 33	87.2	3.8	852	14	US-10-198-846-255	Sequence 255, App
34	77	3.4	527	14	US-10-066-543-2156	Sequence 2156, App
35	75.4	3.3	785	14	US-10-102-806-325	Sequence 325, App
36	75.4	3.3	2117	10	US-09-965-529-53	Sequence 53, Appl
37	75.4	3.3	2117	11	US-09-969-680A-53	Sequence 53, Appl
38	67.4	3.0	2317	14	US-10-097-340-94	Sequence 94, Appl
39	62.2	2.7	574	9	US-09-864-761-228	Sequence 228, App
40	62.2	2.7	669	9	US-09-864-761-17051	Sequence 17051, A
c 41	60.2	2.6	60	12	US-10-017-161-2179	Sequence 2179, App
42	60	2.6	60	12	US-09-908-975-15856	Sequence 15856, A
c 43	59.8	2.6	285	9	US-09-737-178-64	Sequence 64, Appl
c 44	59.8	2.6	285	10	US-09-286-488-64	Sequence 64, Appl
c 45	59.8	2.6	285	11	US-09-853-079-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1

US-10-171-581-170
 ; Sequence 170, Application US/10171581
 ; Publication No. US20030104426A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: Linsley, Peter
 ; APPLICANT: Mao Mao
 ; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
 ; FILE REFERENCE: 9301-157-999
 ; CURRENT APPLICATION NUMBER: US/10/171,581
 ; CURRENT FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: 60/298,914
 ; PRIOR FILING DATE: 2001-06-18
 ; NUMBER OF SEQ ID NOS: 366
 ; SEQ ID NO 170
 ; LENGTH: 2744
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: U41060
 ; DATABASE ENTRY DATE: 2001-06-18
 US-10-171-581-170

Query Match	98.0%	Score 2222;	DB 14;	Length 2744;
Best Local Similarity	99.2%;	Pred. No. 0;		
Matches 2250;	Conservative	0;	Mismatches	0;
			Indels	18;
			Gaps	1;
Qy	1	ATGGCGAGGAGTATCTGTATCTTGTATCTGACCTTTCGCCCTTCCTGTCACAAATCCC	60	
Db	138	ATGGCGAGGAGTATCTGTATCTTGTATCTGACCTTTCGCCCTTCCTGTCACAAATCCC	197	
Qy	61	CTTATGACTAATTAAGCAGCTGCTTTCGCCAGACCTGAGAAATTTAGTCCGAAATGG	120	
Db	198	CTTATGACTAATTAAGCAGCTGCTTTCGCCAGACCTGAGAAATTTAGTCCGAAATGG	257	
Qy	121	GAATCTGCTAATTTGTGACTTTGGCAATTTCCACAGGCAATATCATCTACACAGCTT	180	
Db	258	GAATCTGCTAATTTGTGACTTTGGCAATTTCCACAGGCAATATCATCTACACAGCTT	317	

QY	181	TTCTACGGCTATGGAGAAAATAATTCCTTGTCCAGTTGAAGGGTTTCAGAAAATTACTTCAA	240
Db	318	TTCTACGGCTATGGAGAAAATAATTCCTTGTCCAGTTGAAGGGTTTCAGAAAATTACTTCAA	377
QY	241	AATATAGGCATATGATTAAGTTAAAGAAATCCCATATACACCATGACCAACGACCATCACTCA	300
Db	378	ATATATAGGCATATGATTAAGTTAAAGAAATCCCATATACACCATGACCAACGACCATCACTCA	437
QY	301	GACCACAGAGTCATCTCAGACCATGAGCGGTCACTCAGACCATGAGCATCACTTCAGACAC	360
Db	438	GACCACAGAGTCATCTCAGACCATGAGCGGTCACTCAGACCATGAGCATCACTTCAGACAC	497
QY	361	GAGCATCACTCTGTCACCATGATCATCACTCTCACCATAATCATGCTGCTTCTGGTAAAAAT	420
Db	498	GAGCATCACTCTG-----ACCATATCATGCTGCTTCTGGTAAAAAT	539
QY	421	AAGCGAAAAGCTCTTTGGCCCGACGACCATGACTCAGATAGTTTCAGGTAAAGATCCTTAGAAC	480
Db	540	AAGCGAAAAGCTCTTTGGCCCGACGACCATGACTCAGATAGTTTCAGGTAAAGATCCTTAGAAC	599
QY	481	AGCCACGGGGAAGGAGCTCACCCACGACGAACTGCCAGTGGTGAAGGAATGTCGAAGGAC	540
Db	600	AGCCACGGGGAAGGAGCTCACCCACGACGAACTGCCAGTGGTGAAGGAATGTCGAAGGAC	659
QY	541	AGTGTATAGTCTAGTGAAGTGACCTCACTGTGTACAACTGTCTCTGAAGGAACCTCAC	600
Db	660	AGTGTATAGTCTAGTGAAGTGACCTCACTGTGTACAACTGTCTCTGAAGGAACCTCAC	719
QY	601	TTTCTAGAGACAATAGAGACATCCAGACCTCGAAAAAACHTCTCCCGCAAGANTGTAGACGC	660
Db	720	TTTCTAGAGACAATAGAGACATCCAGACCTCGAAAAAACHTCTCCCGCAAGANTGTAGACGC	779
QY	661	TCCACTCCACCCAGTGTCACATCAAGAGACCGGGGTGAGCGCGTGGCTGAGGAAAACA	720
Db	780	TCCACTCCACCCAGTGTCACATCAAGAGACCGGGGTGAGCGCGTGGCTGAGGAAAACA	839
QY	721	AATGAATCTGTGAGTAGCGCCCGGAAAAAGGCTTTATGTATTCAGAGAACACAAATGAAAT	780
Db	840	ATAGATCTGTGAGTAGCGCCCGGAAAAAGGCTTTATGTATTCAGAGAACACAAATGAAAT	899
QY	781	CCTCAGGAGTGTTCATATGCATCAAGACTACTGACATCTCATGGCATGGGCATCCAGATT	840
Db	900	CCTCAGGAGTGTTCATATGCATCAAGACTACTGACATCTCATGGCATGGGCATCCAGATT	959
QY	841	CCGCTGAATGCACACAGAGTTCACATCTCTGTCGACGCCATCATCACCAATTATGATGCT	900
Db	960	CCGCTGAATGCACACAGAGTTCACATCTCTGTCGACGCCATCATCACCAATTATGATGCT	1019
QY	901	AGATCTTGTCTGATTCATACAGTAGAAAAGAGGCTGAAATCCCTCCAAAGACCTATTCA	960
Db	1020	AGATCTTGTCTGATTCATACAGTAGAAAAGAGGCTGAAATCCCTCCAAAGACCTATTCA	1079
QY	961	TTACAAATAGCTCGGGTGGTGGTTTATATAGCCATTCATCATCACTGTTCTGCTCTCTG	1020
Db	1080	TTACAAATAGCTCGGGTGGTGGTTTATATAGCCATTCATCATCACTGTTCTGCTCTCTG	1139
QY	1021	CTGGGGGTATCTTAGTGCCTCTCATGAATCGGGTGTTTTTCAAATTTCTCTGAGTTTC	1080
Db	1140	CTGGGGGTATCTTAGTGCCTCTCATGAATCGGGTGTTTTTCAAATTTCTCTGAGTTTC	1199
QY	1081	CTTGTGGCACTGGCGTGTGGCACTTTGAGTGGTGATGCTTTTTTACACCTTTCTTCCACAT	1140
Db	1200	CTTGTGGCACTGGCGTGTGGCACTTTGAGTGGTGATGCTTTTTTACACCTTTCTTCCACAT	1259
QY	1141	TCCTCATCGAAGTCACCCACCATAGTCATPAGCCATGAAGAACCCAGCAATGGAATGAAAAGA	1200
Db	1260	TCCTCATCGAAGTCACCCACCATAGTCATPAGCCATGAAGAACCCAGCAATGGAATGAAAAGA	1319
QY	1201	GGACCACTTTTCAGTCATCTGTCTTCTCAAAACATAGAGAAGAGTGCCTATTTTGAATCC	1260
Db	1320	GGACCACTTTTCAGTCATCTGTCTTCTCAAAACATAGAGAAGAGTGCCTATTTTGAATCC	1379

1261	QY	ACGTGGAGGGTCTAACAGCTCTAGGAGCGCTGTATTTCATGTTTCTGTGTGAACATGTC	1321
1380	DB	ACGTGGAGGGTCTAACAGCTCTAGGAGCGCTGTATTTCATGTTTCTGTGTGAACATGTC	1439
1321	QY	CTCATTGTGATCAAAACAATTTAAAGATAAGAGAAAGAAAGAAATCAGAGAAACCTGAAAT	1380
1440	DB	CTCATTGTGATCAAAACAATTTAAAGATAAGAGAAAGAAAGAAATCAGAGAAACCTGAAAT	1499
1381	QY	GATGATGATGGAGATTAAGAACGAGTTGTCCAGTAGTAATCTCAACTTTCAACAAT	1440
1500	DB	GATGATGATGGAGATTAAGAACGAGTTGTCCAGTAGTAATCTCAACTTTCAACAAT	1559
1441	QY	GAGGAGAAATAGATACAGATGATCGAACTTGAAGGCTATTTACGAGCAGACTCAAGAAG	1500
1560	DB	GAGGAGAAATAGATACAGATGATCGAACTTGAAGGCTATTTACGAGCAGACTCAAGAAG	1619
1501	QY	CCCTCCCACTTTGATPCTCAGACCCGTGAGCTTGGAGAGAGAGAGGTCAATGACCT	1560
1620	DB	CCCTCCCACTTTGATPCTCAGACCCGTGAGCTTGGAGAGAGAGAGGTCAATGACCT	1679
1561	QY	CATGCTCATCCACGAGAGTCTCAATAGTAATGTACCCAGAGGGTGCAGATAAATGC	1620
1680	DB	CATGCTCATCCACGAGAGTCTCAATAGTAATGTACCCAGAGGGTGCAGATAAATGC	1739
1621	QY	CAITTCATTTCCAGATACACTCGGCCACTGACGAGTCTCATTCACACCATCATGAC	1680
1740	DB	CATTTCATTTCCAGATACACTCGGCCACTGACGAGTCTCATTCACACCATCATGAC	1799
1681	QY	TACCATCATTTCTCCATCATCACACCCACCAAAACACCATCTCTCAGTCCACAGCCAG	1740
1800	DB	TACCATCATTTCTCCATCATCACACCCACCAAAACACCATCTCTCAGTCCACAGCCAG	1859
1741	QY	CGCTACTCTCGGAGAGAGCTGAAGAATGCGCGCGTGCACACTTTGGCCTGGATGGTATA	1800
1860	DB	CGCTACTCTCGGAGAGAGCTGAAGAATGCGCGCGTGCACACTTTGGCCTGGATGGTATA	1919
1801	QY	ATGGGTGATGGCTGCACAAATTCAGAGATGGCTAGCAAAATGTGCTGCTTTTACTGAA	1860
1920	DB	ATGGGTGATGGCTGCACAAATTCAGAGATGGCTAGCAAAATGTGCTGCTTTTACTGAA	1979
1861	QY	GCCTATCAAGTGGTTTAAGTACTTCTTTGCTGTGTGTTCTGTCTATGAGTTGCCCTATGAA	1920
1980	DB	GCCTATCAAGTGGTTTAAGTACTTCTTTGCTGTGTGTTCTGTCTATGAGTTGCCCTATGAA	2039
1921	QY	TTAGTGTACTTTGCTTCTACTAAAGGCTGGCATGACCGTTAAAGAGGCTGTCCTTTAT	1980
2040	DB	TTAGTGTACTTTGCTTCTACTAAAGGCTGGCATGACCGTTAAAGAGGCTGTCCTTTAT	2099
1981	QY	AATGCATGTGCACCATGTGCGTATCTTTGAGTAGGCAACAGGAATTTTCATGGTCAAT	2040
2100	DB	AATGCATGTGCACCATGTGCGTATCTTTGAGTAGGCAACAGGAATTTTCATGGTCAAT	2159
2041	QY	TATGTGAAATGTTTCTATGTGGATATTTGCATTTACTGCTGGCTTATTCATGATGTT	2100
2160	DB	TATGTGAAATGTTTCTATGTGGATATTTGCATTTACTGCTGGCTTATTCATGATGTT	2219
2101	QY	GCTCTGTTGATATGTGTACCTGGAATGCTCCCAATATGCTAGTGACCATGGATGTAGC	2160
2220	DB	GCTCTGTTGATATGTGTACCTGGAATGCTCCCAATATGCTAGTGACCATGGATGTAGC	2279
2161	QY	CGCTGGGGATTTCTTTTACAGAACTGCGGATGCTTTTGGGTGTTTGGAAATATGTTA	2220
2280	DB	CGCTGGGGATTTCTTTTACAGAACTGCGGATGCTTTTGGGTGTTTGGAAATATGTTA	2339
2221	QY	CTTATTTCCATATTGGAACATAAATCGTGTTTCGTATTAATTTCTAG	2268
2340	DB	CTTATTTCCATATTGGAACATAAATCGTGTTTCGTATTAATTTCTAG	2387

Db	1560	GAGGAGAAAGTAGATACAGATGATCGAACTGAAGCGTATTTACGACGACGTCAACAGAG	1561	9
QY	1501	CCCTCCCACTTTGATATTCACGACGCTGCAAGTCTTTGGAAGAAGAAAGAGGTCAATGAGCT	1502	1560
Db	1620	CCCTCCCACTTTGATATTCACGACGCTGCAAGTCTTTGGAAGAAGAAAGAGGTCAATGAGCT	1621	1679
QY	1561	CATGCTCATCCACAGGAAGTCTCAATGAATATGATPCCAGAGGGTGCAGAATAAATGTC	1562	1620
Db	1680	CATGCTCATCCACAGGAAGTCTCAATGAATATGATPCCAGAGGGTGCAGAATAAATGTC	1681	1739
QY	1621	CATTACATTTCCACGATACACTGGCCAGTCCAGACGATCTCATTTACACCAACATCATGAC	1622	1680
Db	1740	CATTACATTTCCACGATACACTGGCCAGTCCAGACGATCTCATTTACACCAACATCATGAC	1741	1799
QY	1681	TACCATCATATTTCCATCATCACACACCAACCAACCACTCCTCACAGTCCACAGCCAG	1682	1740
Db	1800	TACCATCATATTTCCATCATCACACCAACCAACCACTCCTCACAGTCCACAGCCAG	1801	1859
QY	1741	CGTACTCTCGGAGGAGCTGAAAGATGCGCGCGPCGCCAATTTGGCTCGATGGTGATA	1742	1800
Db	1860	CGTACTCTCGGAGGAGCTGAAAGATGCGCGCGPCGCCAATTTGGCTCGATGGTGATA	1861	1919
QY	1801	ATGGGTGATGGCCTGCACAAATTCAGGGATGGCCTAGCAATTTGGTCTCTTTACTGAA	1802	1860
Db	1920	ATGGGTGATGGCCTGCACAAATTCAGGGATGGCCTAGCAATTTGGTCTCTTTACTGAA	1921	1979
QY	1861	GGCTTATCAAGTGGTTAAAGTACTTCTGTTGCTGTGTTCTGTCATCAGTGGCTCATGAA	1862	1920
Db	1980	GGCTTATCAAGTGGTTAAAGTACTTCTGTTGCTGTGTTCTGTCATCAGTGGCTCATGAA	1981	2039
QY	1921	TTAGTGTACTTTGTTCTTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGCTCCTTTAT	1922	1980
Db	2040	TTAGTGTACTTTGTTCTTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGCTCCTTTAT	2041	2099
QY	1981	AATGTCATTTGCAGCCATGCTGGGGTACTTTGGAATGGCACACAGGAATTTTCATTTGGTCAT	1982	2040
Db	2100	AATGTCATTTGCAGCCATGCTGGGGTACTTTGGAATGGCACACAGGAATTTTCATTTGGTCAT	2101	2159
QY	2041	TATGCTGAAAATGTTTCTATGTGGATATTTGACATTTACTGCTGGCTTATTCATGATGTT	2042	2100
Db	2160	TATGCTGAAAATGTTTCTATGTGGATATTTGACATTTACTGCTGGCTTATTCATGATGTT	2161	2219
QY	2101	GCCTCGGTTGATGGTAACTGGAATTCGTCACAAATGATGCTAGTGACCATGGATGAGC	2102	2160
Db	2220	GCCTCGGTTGATGGTAACTGGAATTCGTCACAAATGATGCTAGTGACCATGGATGAGC	2221	2279
QY	2161	CGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTTCGAATTTATGTTA	2162	2220
Db	2280	CGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTTCGAATTTATGTTA	2281	2339
QY	2221	CTTATTTCCATATTTGACATAAAATCGTTTTCGTATFAAATTTCTAG	2222	2368
Db	2340	CTTATTTCCATATTTGACATAAAATCGTTTTCGTATFAAATTTCTAG	2341	2387

RESULT 3

```

US-10-198-846-10007
; Sequence 10007, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084

```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 10007
;
; LENGTH: 3537
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: misc_feature
;
; LOCATION: 1, 2, 3533, 3534, 3535, 3536, 3537
;
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10007

```

Query Match	97.9%; Score 2220.4; DB 14; Length 3537;
Best Local Similarity	99.2%; Fred. No. 0;
Matches 2249; Conservative	0; Mismatches 1; Indels 18; Gaps 1;
Qy	1 ATGCGAGGAAGTTATCTGTAAATCTTCATCCTTGACCTTGGCCCTCTCTGTGCACAATCCC 60
Db	138 ATGCGAGGAAGTTATCTGTAAATCTTCATCCTTGACCTTGGCCCTCTCTGTGCACAATCCC 197
Qy	61 CTTCATGAATTAAGCAAGCTGCTTTCCCCAGACCACTGAGAAAAATTAGTCGAAATGG 120
Db	198 CTTCATGAATTAAGCAAGCTGCTTTCCCCAGACCACTGAGAAAAATTAGTCGAAATGG 257
Qy	121 GAATCTGGCAATTAAGTTGACTTGGCAATTTCCACACGGCAATATCATCTACACAGCTT 180
Db	258 GAATCTGGCAATTAAGTTGACTTGGCAATTTCCACACGGCAATATCATCTACACAGCTT 317
Qy	181 TTCACCGCTATGAGAGAAAATAATCTTTGTCACTGAAGGGTTTCAGAAAATTTACTCAA 240
Db	318 TTCTACCGCTATGAGAGAAAATAATCTTTGTCACTGAAGGGTTTCAGAAAATTTACTCAA 377
Qy	241 AATATAGGCATAGATAGATTAAAGAATCCATPATACACCATGACCACGACCACTACATCA 300
Db	378 AATATAGGCATAGATAGATTAAAGAATCCATPATACACCATGACCACGACCACTACATCA 437
Qy	301 GACCACGAGCATCACTCAGACCACTGACGGTCACTCAGACCATGAGCATCACTCAGACCA 360
Db	438 GACCACGAGCATCACTCAGACCACTGACGGTCACTCAGACCATGAGCATCACTCAGACCA 497
Qy	361 GAGCATCACTCTGACCATGATCACTCTCACCAATATCATGCTTGGCTTCTGTAATAAAT 420
Db	498 GAGCATCACTCTG-----ACCATATCATGCTGCTTCTGGTAAAAAT 539
Qy	421 AAGCGAAAGCTCTTTGCCAGACCATGACTCAGATGATTCAGTAAAGATCTTAGAAC 480
Db	540 AAGCGAAAGCTCTTTGCCAGACCATGACTCAGATGATTCAGTAAAGATCTTAGAAC 599
Qy	481 AGCCAGGGGAAGAGAGCTCACCGACCAAGAACATGCGAGTGTAGAAGGAATGTCAAGAC 540
Db	600 AGCCAGGGGAAGAGAGCTCACCGACCAAGAACATGCCAGTGTAGAAGGAATGTCAAGAC 659
Qy	541 AGTGTATGCTAGTAGAGTGACCTCAACTGHTGTACAACTGTCTCTGAAGGAATCTAC 600
Db	660 AGTGTATGCTAGTAGAGTGACCTCAACTGHTGTACAACTGTCTCTGAAGGAATCTAC 719
Qy	601 TTTCTAGAGACATAGAGACTCCAAGACCTGSGAAACCTTCCCCRAAGATGTAAAGCAGC 660
Db	720 TTTCTAGAGACATAGAGACTCCAAGACCTGSGAAACCTTCCCCRAAGATGTAAAGCAGC 779
Qy	661 TCCACTCCACCCAGTGTCCACATCAAGAGCGCGGGTGAAGCGGCTGGCTGGTAGGAAAA 720
Db	780 TCCACTCCACCCAGTGTCCACATCAAGAGCGCGGGTGAAGCGGCTGGCTGGTAGGAAAA 839
Qy	721 AATGAATCTGTAGTGAGCCCGGAAAAAGGCTTTATGTATTCAGAAACACAAATGAAAT 780
Db	840 AATGAATCTGTAGTGAGCCCGGAAAAAGGCTTTATGTATTCAGAAACACAAATGAAAT 899
Qy	781 CCTCAGAGTGTTTTCAATGCATCAAGCTACTGACATCTCATGGCATGGCATCCAGGTT 840
Db	900 CCTCAGAGTGTTTTCAATGCATCAAGCTACTGACATCTCATGGCATGGCATCCAGGTT 959
Qy	841 CCGCTGAATGCACAGAGTTTCAACTATCTCTGTCGAGGCATCATCAACCAAAATTTGATGCT 900

Db 960 CCGCTGAATGCAACAGAGCTTCACTATCTCTGTCCAGCCATCATCAACCAAAATGATGCT 1019
QY 901 AGATCTTGCTGATTCATACAAAGTGAAAGAGGCTGAAATCCCTCCAAAGACCTATTCA 960
Db 1020 AGATCTTGCTGATTCATACAAAGTGAAAGAGGCTGAAATCCCTCCAAAGACCTATTCA 1079
QY 961 TTACAAATAGCCCTGGGTGGTGGTTTATAGCCATTTCCATCATCAGTTTCCGTGCTCG 1020
Db 1080 TTACAAATAGCCCTGGGTGGTGGTTTATAGCCATTTCCATCATCAGTTTCCGTGCTCG 1139
QY 1021 CTGGGGTATCTTAGTGCCCTCATGAATCGGGTGGTTTCAAAATTTCCGAGTTTC 1080
Db 1140 CTGGGGTATCTTAGTGCCCTCATGAATCGGGTGGTTTCAAAATTTCCGAGTTTC 1199
QY 1081 CTGTGGCACTGECCTGTTGGACATTTGAGTGGTGATGCTTTTACACCTTCTTCCACAT 1140
Db 1200 CTGTGGCACTGECCTGTTGGACATTTGAGTGGTGATGCTTTTACACCTTCTTCCACAT 1259
QY 1141 TCTCATGCAAGTCACCACTAGTCATGAGTGGTGATGCTTTTACACCTTCTTCCACAT 1200
Db 1260 TCTCATGCAAGTCACCACTAGTCATGAGTGGTGATGCTTTTACACCTTCTTCCACAT 1319
QY 1201 GGACCACTTTTCAGTCATCTCTCTTCAAAACATAGAGAAAGTGCTATTTTGATTC 1260
Db 1320 GGACCACTTTTCAGTCATCTCTCTTCAAAACATAGAGAAAGTGCTATTTTGATTC 1379
QY 1261 ACSTGGNAGGCTTACAGCTCTAGAGGCTGTATTTTCATGTTTCTTGTGACATGTC 1320
Db 1380 ACSTGGNAGGCTTACAGCTCTAGAGGCTGTATTTTCATGTTTCTTGTGACATGTC 1439
QY 1321 CTCACATGATCAAACTTTAAAGATAAGAGAAAGAAATCAGAGAACTGAAAT 1380
Db 1440 CTCACATGATCAAACTTTAAAGATAAGAGAAAGAAATCAGAGAACTGAAAT 1499
QY 1381 GATGATGATGAGATGAAGACGATTTCCAAATATGATGATCTCACTTCAACAAAT 1440
Db 1500 GATGATGATGAGATGAAGACGATTTCCAAATATGATGATCTCACTTCAACAAAT 1559
QY 1441 GAGGAGAAAGTATACAGATGATCGACATGAGGCTATTTACGAGCAGCTCACAGAG 1500
Db 1560 GAGGAGAAAGTATACAGATGATCGACATGAGGCTATTTACGAGCAGCTCACAGAG 1619
QY 1501 CCTCCCACTTTGATTCACAGCCTGCTGCTTGGAGAGAGAGGTCATGATGCT 1560
Db 1620 CCTCCCACTTTGATTCACAGCCTGCTGCTTGGAGAGAGAGGTCATGATGCT 1679
QY 1561 CATGCTATCCACAGAAAGTCTACAATGAATATGATGATGATGATGATGATGATGATG 1620
Db 1680 CATGCTATCCACAGAAAGTCTACAATGAATATGATGATGATGATGATGATGATGATG 1739
QY 1621 CATTCACATTTCCACAGATACACTCGGCAGTCAGAGATCTCATTCACCACTCATGAC 1680
Db 1740 CATTCACATTTCCACAGATACACTCGGCAGTCAGAGATCTCATTCACCACTCATGAC 1799
QY 1681 TACCATCATTTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1740
Db 1800 TACCATCATTTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1859
QY 1741 CGCTACTCTCGGAGAGGCTGAAAGATGCGCGGCTCGCCACTTGGCCCTGGATGGTGA 1800
Db 1860 CGCTACTCTCGGAGAGGCTGAAAGATGCGCGGCTCGCCACTTGGCCCTGGATGGTGA 1919
QY 1801 ATGGGTGATGGCTGCACAAATTTACGAGATGGCTAGCAATTTGGTCTTTTACTGAA 1860
Db 1920 ATGGGTGATGGCTGCACAAATTTACGAGATGGCTAGCAATTTGGTCTTTTACTGAA 1979
QY 1861 GGCTTATCAAGTGGTAAAGTACTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1980 GGCTTATCAAGTGGTAAAGTACTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTTAT 2039
QY 1921 TTAGGTGACTTTTCTGCTTACTAAAGGCTGGCATGACCGTTAAAGCGGCTGCTTAT 1980
Db 2040 TTAGGTGACTTTTCTGCTTACTAAAGGCTGGCATGACCGTTAAAGCGGCTGCTTAT 2099

QY 1981 AATGCATTTCCAGCATGCTGGGCTATCTTGGAAATGCAACAGAAATTTTCATTTGTCAT 2040
Db 2100 AATGCATTTCCAGCATGCTGGGCTATCTTGGAAATGCAACAGAAATTTTCATTTGTCAT 2159
QY 2041 TATGCTGAAATGTTTCTATGTGGATATTTGCACCTTACTGCTGCTTATTCATGATGTT 2100
Db 2160 TATGCTGAAATGTTTCTATGTGGATATTTGCACCTTACTGCTGCTTATTCATGATGTT 2219
QY 2101 GCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160
Db 2220 GCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2279
QY 2161 CGTGGGGTATCTTTTACAGATGCTGGGATGCTTTGGGTTTGGAAATATGTTA 2220
Db 2280 CGTGGGGTATCTTTTACAGATGCTGGGATGCTTTGGGTTTGGAAATATGTTA 2339
QY 2221 CTATTTCCATATTTGAACTAAATCGTGTTCGTATATAATTTCTAG 2268
Db 2340 CTATTTCCATATTTGAACTAAATCGTGTTCGTATATAATTTCTAG 2387

RESULT 4
US-10-176-847-51
; Sequence 51, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Velby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREA
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREA
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 3586
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-51

Query Match 96.9%; Score 2198; DB 14; Length 3586;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2248; Conservative 0; Mismatches 0; Indels 20; Gaps 3;

QY 1 ATGCGGAGAAATATCTGTAATCTTGATGCTGCTTGGCCCTCTCTGTCAACAATCC 60
Db 263 ATGCGGAGAAATATCTGTAATCTTGATGCTGCTTGGCCCTCTCTGTCAACAATCC 322
QY 61 CTTTCATGAACCTAAAGCAGCTGCTTCCCGCAGACCTGAGAAATTAGTCCGAATGG 120
Db 323 CTTTCATGAACCTAAAGCAGCTGCTTCCCGCAGACCTGAGAAATTAGTCCGAATGG 382
QY 121 GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACGCGCAATATCATCTACACAGCTT 180
Db 383 GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACGCGCAATATCATCTACACAGCTT 442
QY 181 TTCTACCGGTATGAGAGAAATTAATCTTTGTCAGTTGAGGGTTTCAGAAAATTTACTTCA 240
Db 443 TTCTACCGGTATGAGAGAAATTAATCTTTGTCAGTTGAGGGTTTCAGAAAATTTACTTCA 502
QY 241 AATATAGGCATAGATAAGATTAAGAAATCCATATACACCATGACACCATCATCTCA 300
Db 503 AATATAGGCATAGATAAGATTAAGAAATCCATATACACCATGACACCATCATCTCA 562
QY 301 GACCAAGCAGCATCTCAGACCATGAGCGTCTACTCAGACCATGAGCATCTCAGACCAT 360
Db 563 GACCAAGCAGCATCTCAGACCATGAGCGTCTACTCAGACCATGAGCATCTCAGACCAT 622
QY 361 GAGCATCACTCTGACCATGATCATCTCTCACCATATCATGCTGCTTCTGGTAAATAT 420
Db 623 GAGCATCACTCTGACCATGATCATCTCTCACCATATCATGCTGCTTCTGGTAAATAT 664

; ORGANISM: Homo sapiens
US-09-925-300-591

```
Query Match          50.3%; Score 1141.4; DB 10; Length 1193;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1139; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 539 ACAGTGTAGTCTAGTGAAGTGAAGTCAACTGTGTACAACTGTCTCTGAAGGAATC 598
Db 1 ACAGTGTAGTCTAGTGAAGTGAAGTCAACTGTGTACAACTGTCTCTGAAGGAATC 60

QY 599 ACTTTCAGACATAGACATCCAGACCTGGAAGAACTCTCCCAAGAGTGAAGCA 658
Db 61 ACTTTCAGACATAGACATCCAGACCTGGAAGAACTCTCCCAAGAGTGAAGCA 120

QY 659 GCTCCACTCCACCAGTGTACATCAAGAGCGGGTGGCGGTGGTAGGAAA 718
Db 121 GCTCCACTCCACCAGTGTACATCAAGAGCGGGTGGCGGTGGTAGGAAA 180

QY 719 CAATGAATCTGTAGTGAGCCCGGAAAGGCTTTATGTATTCAGAGAAACAAATGAAA 778
Db 181 CAATGAATCTGTAGTGAGCCCGGAAAGGCTTTATGTATTCAGAGAAACAAATGAAA 240

QY 779 ATCCTCAGAGTGTTCATCATCAAGCTACTGACATCTCATGGCATGGCATCCAG 838
Db 241 ATCCTCAGAGTGTTCATCATCAAGCTACTGACATCTCATGGCATGGCATCCAG 300

QY 839 TTCGGTGAATGCAACAGAGTCAACTPATCTGTCCAGCCATCATCAACAAATGATG 898
Db 301 TTCGGTGAATGCAACAGAGTCAACTPATCTGTCCAGCCATCATCAACAAATGATG 360

QY 899 CTAGACTTGTCTGATTCATACAGTGAAGAGGCTGAATCCCTCCAAAGACCTATT 958
Db 361 CTAGACTTGTCTGATTCATACAGTGAAGAGGCTGAATCCCTCCAAAGACCTATT 420

QY 959 CATTCAAAATAGCTGGTGGTGTATAGCCATTTCATCATCAGTTTCCCTGTC 1018
Db 421 CATTCAAAATAGCTGGTGGTGTATAGCCATTTCATCATCAGTTTCCCTGTC 480

QY 1019 TGTGGGGTGTATCTATGCTCTCATGATGAGGCTTTTCAAAATTTCTCCTGAGTT 1078
Db 481 TGTGGGGTGTATCTATGCTCTCATGATGAGGCTTTTCAAAATTTCTCCTGART 540

QY 1079 TCCTTGTGCACTGCGCGTGGGACTTTGAGTGGTGATGCTTTTACACTTCTCCAC 1138
Db 541 YCCTTGTGCACTGCGCGTGGGACTTTGAGTGGTGATGCTTTTACACTTCTCCAC 600

QY 1139 ATTCTCATCAAGTCAACCAATAGTCATAGCCATGAAGAACCAAGCAATGAAATGAAA 1198
Db 601 ATTCTCATCAAGTCAACCAATAGTCATAGCCATGAAGAACCAAGCAATGAAATGAAA 660

QY 1199 GAGGACACTTTTCAGTATCTGCTCTCAAAACATAGAGAAAGTGCCTATTTTGAT 1258
Db 661 GAGGACACTTTTCAGTATCTGCTCTCAAAACATAGAGAAAGTGCCTATTTTGAT 720

QY 1259 CCAGTGGAGGGTCTAAAGCTCTAGAGGCGCTGATTTTCATGTTTCTGTTGAACATG 1318
Db 721 CCAGTGGAGGGTCTAAAGCTCTAGAGGCGCTGATTTTCATGTTTCTGTTGAACATG 780

QY 1319 TCCTCAGATGTATCAAAATTTAAAGATGAAGAAAGAAAGATCAGAGAAACCTGAAA 1378
Db 781 TCCTCAGATGTATCAAAATTTAAAGATGAAGAAAGAAAGATCAGAGAAACCTGAAA 840

QY 1379 ATGATGATGATGTGAGATTAAGAGCAGTTGCTCCAGTATGATCTCACTTTCAACAA 1438
Db 841 ATGATGATGATGTGAGATTAAGAGCAGTTGCTCCAGTATGATCTCACTTTCAACAA 900

QY 1439 ATGAGGAGAAAGTATGATACATGATCGAACTGAAGGCTATTATGAGCAGACTCAAG 1498
Db 901 ATGAGGAGAAAGTATGATACATGATCGAACTGAAGGCTATTATGAGCAGACTCAAG 960

QY 1499 AGCCCTCCACTTTGATTTCTCAGCAGGCTCGACTGTGGAAGAGAAAGGTGATGAT 1558
Db 1558 AGCCCTCCACTTTGATTTCTCAGCAGGCTCGACTGTGGAAGAGAAAGGTGATGAT
```

```
Db 961 AGCCCTCCACTTTGATTTCTCAGCAGCCTGCAGTCTTGGAAAGAAAGAGGTCATGATAG 1020
QY 1559 CTCATGCTCATCCACAGGAAGTCTACAATGAATATGTATCCAGAGGGTGCAGAAATTAAT 1618
Db 1021 CTCATGCTCATCCACAGGAAGTCTACAATGAATATGTATCCAGAGGGTGCAGAAATTAAT 1080
QY 1619 GCCATTCATATTCACAGATACACTCGGCCAGTCAGAGATCTCAATTCACACCATCATG 1678
Db 1081 GCCATTCATATTCACAGATACACTCGGCCAGTCAGAGATCTCAATTCACACCATCATG 1140
QY 1679 ACT 1681
Db 1141 ACT 1143

RESULT 6
US-10-198-846-10321
; Sequence 10321, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lilie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steimann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; FILE REFERENCE: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10321
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1447, 1448, 1449
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10321

Query Match          39.7%; Score 900.4; DB 14; Length 1449;
Best Local Similarity 96.4%; Pred. No. 8e-248;
Matches 985; Conservative 0; Mismatches 31; Indels 6; Gaps 6;

QY 967 ATAGCCTGGGTGGTGGTGTATAGCCATTCCATCATCAGTTTCCCTGCTCTGCTGGGG 1026
Db 199 ACAGCCTGGGTGGTGGTGTATAGCCATTCCATCATCAGTTTCCCTGCTCTGCTGGGG 258

QY 1027 GTTATCTTAGTGCCTCTCATGAATCGGGGTGTTTTCAAATTTCCCTGAGTTTCCCTGTG 1086
Db 259 GTTATCTTAGTGCCTCTCATGAATCGGGGTGTTTTCAAATTTCCCTGAGTTTCCCTGTG 318

QY 1087 GCATCGCGGTGGGACTTTGAGTGGTGATGCTTTTACACCTCTTCCACATTCAT 1146
Db 319 GCATCGCGGTGGGACTTTGAGTGGTGATGCTTTTACACCTCTTCCACATTCAT 378

QY 1147 GCAAGTCACCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1206
Db 379 GCAAGTCACCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 438

QY 1207 CTTTTCAGTCACTCTCTCTCTCAAAACATGAGAAAGTGCCTATTTTGTATTCACGTGG 1266
Db 439 CTTTTCAGTCACTCTCTCTCTCAAAACATGAGAAAGTGCCTATTTTGTATTCACGTGG 498

QY 1267 AAGGCTCAACAGCTCTAGAGGCGCTGATTTTCATGTTTCTTGTGAACATGCTCTACA 1326
Db 499 AAGGCTCAACAGCTCTAGAGGCGCTGATTTTCATGTTTCTTGTGAACATGCTCTACA 558

QY 1327 TTGATCAAAACAATTTAAAGATAAGAAAGAAAGAAAGTCAAGAAACCTGAAATGATGAT 1386
```



```
|||||
Db 559 TTGATCAACAAATTAAGATAGAGAGAAAAGATCAGAGAAACCTGAAATGATGAT 618
Qy 1387 GATGTGGAGATTAAAGAGAGTGTGCCAAGTATGAATCTCAACTTTCAACAATGAGGAG 1446
Db 619 GATGTGGAGATTAAAGAGAGTGTGCCAAGTATGAATCTCAACTTTCAACAATGAGGAG 678
Qy 1447 AAAGTAGATACAGATGATCGAATGAGGCTATTAGAGACACTCAGAGAGCCCTCC 1506
Db 679 AAAGTAGATACAGATGATCGAATGAGGCTATTAGAGACACTCAGAGAGCCCTCC 738
Qy 1507 CACTTTGATTTCCAGAGCCTGCAGTCTTTGGAAGAAGAGGTCATGATAGCTCANGCT 1566
Db 739 CACTTTGATTTCCAGAGCCTGCAGTCTTTGGAAGAAGAGGTCATGATAGCTCANGCT 798
Qy 1567 CATCCACAGGAAGTCTCAAGTAATATGTAACAGAGGTCGAAGTAATATGCAATCA 1626
Db 799 CATCCACAGGAAGTCTCAAGTAATATGTAACAGAGGTCGAAGTAATATGCAATCA 858
Qy 1627 CATTTCCAGATACACTCGGCCAGTCAGAGAGTCTCATTCACACCATCATGACTACCAT 1686
Db 859 CATTTCCAGATACACTCGGCCAGTCAGAGAGTCTCATTCACACCATCATGACTACCAT 918
Qy 1687 CATATTTCCATCATCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1745
Db 919 CATATTTCCATCATCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 978
Qy 1746 CTCCTGGAGAGAGTGAAGATCGCGCGTGCACACTTTGGCTGGATGGTGAATGAGG 1805
Db 979 CTCCTGGAGAGAGTGAAGATCGCGCGTGCACACTTTGGCTGGATGGTGAATGAGG 1038
Qy 1806 TGATGGCTCGACAA--TTTCAGCGATGGCTAGCAATGG--TGCTGCTTTTACTGAA--GG 1862
Db 1039 TGATGGCTCGACAA--TTTCAGCGATGGCTAGCAATGG--TGCTGCTTTTACTGAA--GG 1098
Qy 1863 CTTATCAAGTGGTTAAGTACTTCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 1922
Db 1099 CTTATCAAGTGGTTAAGTACTTCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 1157
Qy 1923 AGTGAGCTTTCCTGTTCTACTAAAGGCTGGATGACCGTTAAGCAGGCTGCTCTTATA 1982
Db 1158 AGAGACTAAGCTGTTCTACTACAGGAGGAGGCAATACCG--TAAGCCGGCTGCTCTTGCA 1216
Qy 1983 TG 1984
Db 1217 TG 1218
```

```
RESULT 7
US-198-846-13805
; Sequence 13805, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIORITY FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13805
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013,
; LOCATION: 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023,
; LOCATION: 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-13805

Query Match 30.6%; Score 694.2; DB 14; Length 1032;
Best Local Similarity 90.0%; Pred. No. 1.3e-188;
Matches 800; Conservative 0; Mismatches 83; Indels 6; Gaps 5;

Qy 214 GTTGAAGGGTTCAGAAAATTACTTCAAAATATAGGATAGATAAGATTAAAGAATCCAT 273
Db 4 GTCCGAGGGTTAGAAAATTACTT-AAAATATAGGATAGATAAGATTAAAGAATCCAT 62
Qy 274 ATACACATGACACACGACCATCATCTCAGACAGAGCATCAC-TCAGACCATGAGC-6TC 331
Db 63 ATACACATGACACACGACCATCATCTCAGACAGAGCATCACGTCAGACCATGAGCGGTC 122
Qy 332 ACTCAGACCATGAGC-ATCAGTCAGACAGAGCATCTCTGACCATGATCATCTCT 390
Db 123 ACTCAGACCATGAGC-ATCAGTCAGACAGAGCATCTCTGACCATGATCATCTCT 182
Qy 391 CACCATAATCATCTCTCTTGGTAAATAAGCGAAAAGCTCTTTGCCACAGACATGAC 450
Db 183 CACCATAATCATCTCTCTTGGTAAATAAGCGAAAAGCTCTTTGCCACAGACATGAC 242
Qy 451 TCAGATAGTTTCAGGTAAAGATCCTTAGAACAAGCCAGGGAAGAGGCTCACCGACAGAA 510
Db 243 TCAGATAGTTTCAGGTAAAGATCCTTAGAACAAGCCAGGGAAGAGGCTCACCGACAGAA 302
Qy 511 CATGCCAGTGGTAAAGGAAATGTCGAAGGACAGTGTAGTGTAGTGAAGTGAAGTCAACT 570
Db 303 CATGCCAGTGGTAAAGGAAATGTCGAAGGACAGTGTAGTGTAGTGAAGTGAAGTCAACT 362
Qy 571 GGTACAACTCTCTCTCTGAAGAACTCCTCTCTAGAGCAATAGAGACTTCAAGACCT 630
Db 363 GTTACAACTCTCTCTCTGAAGAACTCCTCTCTAGAGCAATAGAGACTTCAAGACCT 422
Qy 631 GGAAGAACTCTCCCAAGATGTAAAGAGTCCACTCCAGGAGTGCACATCAAGAGC 690
Db 423 GGAAGAACTCTCCCAAGATGTAAAGAGTCCACTCCAGGAGTGCACATCAAGAGC 482
Qy 691 CGGGTGAAGCGGCTGGTGGTAGGAAAACAATAAATCTGTGAGTGAAGCCCGAAGAGC 750
Db 483 CGGGTGAAGCGGCTGGTGGTAGGAAAACAATAAATCTGTGAGTGAAGCCCGAAGAGC 542
Qy 751 TTTATGATTTCCAGAAACAATAAATCTGTGAGTGAAGTGTTCATGCTCAAGCTA 810
Db 543 TTTATGATTTCCAGAAACAATAAATCTGTGAGTGAAGTGTTCATGCTCAAGCTA 602
Qy 811 CTGACATCTCATGGCATGGGTCATCCAGGTTCCGCTGAATGCAACAGAGTTCACATCTC 870
Db 603 CTGACATCTCATGGCATGGGTCATCCAGGTTCCGCTGAATGCAACAGAGTTCACATCTC 662
Qy 871 TGTCCAGCATCATCAACCAATTTGATGCTAGATCTTGTCTGATTCATCAAGTGAAG 930
Db 663 TGTCCAGCATCATCAACCAATTTGATGCTAGATCTTGTCTGATTCATCAAGTGAAG 722
Qy 931 AAGGCTGAATTCCTCCAAAGACCTATTCAATACAAATAGGCTGGGTGGTGGTTTATA 990
Db 723 AAGGCTGAATTCCTCCAAAGACCTATTCAATACAAATAGGCTGGGTGGTGGTTTATA 782
Qy 991 GCCATTTCCATCATCAAGTTCCTGCTGCTGGGGTTATCTTAGTGCCTCTCATGAAT 1050
Db 783 GGCATTTCCATCATCAAGTTCCTGCTGCTGGGGTTATCTTAGTGCCTCTCATGAAT 840
Qy 1051 CGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1099
Db 841 CGGGGTTCCACAAATCCCTGGCATTCCTGTGACTGCGCGTGGCACTTG 889
```

```
RESULT 8
US-10-198-846-8490
```

```
; SEQ ID NO 8490
;
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 44, 516, +
; OTHER INFORMATION: n =
```

RESULT 9

```

; LENGTH: 366
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
US-09-918-995-30283

```

309 TGCCCTTTAATGCAATGTCAGCCATGCTGGCGTATCTTGGAAATGCCACAGGAAT 366

ORGANISM:

FEATONE,

```
; NAME/KEY: misc_feature
; LOCATION: 4, 583, 650, 713, 714, 721, 731, 756, 787, 788, 790, 792.
; LOCATION: 800, 802, 804, 811, 816, 832, 838, 840, 844, 854, 859, 873,
; LOCATION: 875, 887, 888, 898
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-7874

Query Match      14.6%; Score 331.8; DB 14; Length 905;
Best Local Similarity 98.0%; Pred. No. 2.2e-84;
Matches 336; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 246 AGGCATAGATAAGATTAAAGATCCATATACACCATGACCCAGCACCATCACTCAGACCA 305
DB 73 AGGCATAGATAAGATTAAAGATCCATATACACCATGACCCAGCACCATCACTCAGACCA 132
QY 306 CGAGCATCACTCAGACCATGAGCTCACTCAGACCATGAGCATCACTCAGACCATGAGCA 365
DB 133 CGAGCATCACTCAGACCATGAGCTCACTCAGACCATGAGCATCACTCAGACCATGAGCA 192
QY 366 TCACCTGACCATGATCATCACTCTCCACCATATCACTGCTCTGCTGTAATATAGCG 425
DB 193 TCACCTGACCATGATCATCACTCTCCACCATATCACTGCTCTGCTGTAATATAGCG 252
QY 426 AAAAGCTCTTTGCCAGACCATGACTCAGATAGTTCAGGTAAAGATCCTAGAAACAGCCA 485
DB 253 AAAAGCTCTTTGCCAGACCATGACTCAGATAGTTCAGGTAAAGATCCTAGAAACAGCCA 312
QY 486 GGGNAAGGAGCTCAGCCAGACATGCGAGTGGGTAGAGGATGTCAAGGACAGTGT 545
DB 313 GGGNAAGGAGCTCAGCCAGACATGCGAGTGGGTAGAGGATGTCAAGGACAGTGT 372
QY 546 TAGTCTAGTGAAGTGAAGTCACTCACTGCTGATCAACACTGTCTCT 588
DB 373 TAGTCTAGTGAAGTGAAGTCACTCACTGCTGATCAACACTGTCTCT 415

RESULT 11
US-09-920-455-250
; Sequence 250, Application US/09920455
; Patent No. US20020168647A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF HEAD AND NECK CANCER
; FILE REFERENCE: 210121.540
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 275
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 42, 43
; OTHER INFORMATION: n = A,T,C or G
US-09-920-455-250

Query Match      14.1%; Score 319.2; DB 10; Length 337;
Best Local Similarity 98.5%; Pred. No. 4.4e-81;
Matches 321; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1333 AAACAATTAAAGATAAGAAAGAAAGATCAGAGAAACCTGAAATGATGATGTG 1392
DB 12 AATTCAATTAAAGATAAGAAAGAAAGATCAGAGAAACCTGAAATGATGATGTG 71
QY 1393 GAGATTAGAGCAGTGTCCCAAGTATGAATCTCAACTTTCACAAATGAGGAGAAAGTA 1452
DB 72 GAGATTAGAGCAGTGTCCCAAGTATGAATCTCAACTTTCACAAATGAGGAGAAAGTA 131
QY 1453 GATACAGATGATGAACTGAAGGCTATTATACGAGCAGACTCACAGAGCCCTCCACTTT 1512
```

```
DB 132 GATACAGATGATCGAAGCTGAAGGCTATTATACGAGCAGACTCACAGAGCCCTCCACTTT 191
QY 1513 GATTCTCAGCAGCCTGCAAGTCTTGGAGAGAGAGAGGTCAATGATGCTCATGCTCATCCA 1572
DB 192 GATTCTCAGCAGCCTGCAAGTCTTGGAGAGAGAGAGGTCAATGATGCTCATGCTCATCCA 251
QY 1573 CAGGAAGTCTCAATGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 1632
DB 252 CAGGAAGTCTCAATGAATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 311
QY 1633 CAGCATACACTCGGCAGCTCAGACGA 1658
DB 312 CAGCATACACTCGGCAGCTCAGACGA 337

RESULT 12
US-09-814-353-22020
; Sequence 22020, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22020
; LENGTH: 4417
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3, 4415, 4416, 4417
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-22020

Query Match      14.1%; Score 319; DB 12; Length 4417;
Best Local Similarity 56.3%; Pred. No. 3.4e-80;
Matches 736; Conservative 0; Mismatches 535; Indels 36; Gaps 6;

QY 969 AGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1028
DB 561 AGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 620
QY 1029 TATCTTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1088
DB 621 GATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 680
QY 1089 ACTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1148
DB 681 ATTAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 740
QY 1149 AAGTCAACCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1208
DB 741 TGGACATGATCAGACGTCACCAACATGACCA 1285
QY 741 TGGACATGATCAGACGTCACCAACATGACCA 1285
```



```
QY 1626 ACATTTCCACGATACACTGGCCAGTCAGAGCATCTCATCCACCATCATGACTACCA 1685
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2066 -CATGAGCATGATCTCTCAATGCTGCGCATCAATCAACACCGCGGAGACAAATCTGCT 2064
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1686 TCATATTTCTCCATCATCAACCCACCAAAACCCACCATCTCATCATCAACGAGGCTA 1745
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2065 GAGGAGCAATATCAACCATGCGCACACAGCATCTCATCATCCATGCGCCCTGTCA 2124
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1746 CTCTCGGAGAGCTGAAGAAGTGGCGGCTGGCCCATCTTTGGCCTGGATGGATATGGG 1805
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2125 TTCTGGATCCGATCTCTGAAGAAGCAAGCAATAGTATATAGCTGGATGGATCATGGG 2184
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1806 TGATGGCTGCACAAATTCAGCGATGGCTAGCAATTTGCTGCTTTTACTGAAGCTT 1865
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2185 GGATGGCATCCACAACTTCAGTATGGGCTCGCAATTTGGTGCAGCTTCAGTCTGGAT 2244
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1866 ATCAAGTGGTTTAAGTACTTCTTGCTGTCTGCTCATGAGTTGCCCTCATGAATFAGG 1925
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2245 GACAGGAGCAATCAGTACTTCTATAGCGTCTTCTGTCATGAAGTGCACATGAATFAGG 2304
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1926 TGACTTGTCTTACTTAAGCTGCGCATGACGCTTAAGCAGGCTGCTCTTATATATGC 1985
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2305 AGATTTTGCAGTTCTTCTTAAAGCAGCATGACTGTAAAGCAAGCAATGTATACACCT 2364
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1986 ATTTGTCAGCATGCTGCGCTATCTTGGAAATGGCAACAGGAATTTTCATTGCTCATATGC 2045
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2365 CTTCTCTGCCATGATGCTTACATAGGCTGCTCATAGGACACAGCTGTTGGTCAATATGC 2424
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2046 TGAATAATGTTTATGTGATATTTGCACTTACTGCTGCTTATTCATGATATGTCTCT 2105
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2425 CAATAATCATCACACTTTGAGTCTTTCAGTCACTGCGGCAATGTTCCCTCTATGAGCTT 2484
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2106 GTTGTATATGTACCTTGAATGCTGCACAAATGATG-----CTAGTGACCATGATATAG 2159
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2485 GGTGATATGCTTCCAGAAATGTTCCATGCTGATGCTGATGATGATGATGATGATG 2544
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2160 CGCTGGGGGATTTCTTTTACAGAAATGCTGGGATGCTTTGGGTTTGGAAATATGTT 2219
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2545 TCGTGTGGGCAATTCATCTTCAGAAATTTAGGATTTGCTTTGGATTTGCCATATGCT 2604
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2220 ACTTATTTCCATATTTGAACATAAATCGTGTTCGTATATAATTTCT 2266
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2605 GGTGATGGCCCTCTATGAAGATAAATTTGTTGTGACATCCAGTTT 2651
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
US-10-198-846-9288/c
; Sequence 9288, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 9288
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8, 11, 42, 262, 279, 324, 333, 337, 344, 346, 349, 352, 374,
; LOCATION: 381, 383, 399, 400, 401, 415, 417, 427, 430, 431, 432, 439,
; LOCATION: 445, 454, 457, 458, 459, 464, 468, 469, 470, 473, 475, 477,
```

```
; LOCATION: 488, 493, 495, 504, 509, 519, 529, 536, 545, 547
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 552, 555, 559, 560, 564, 565, 578, 580, 581, 585, 587, 589,
; LOCATION: 590, 593, 594, 600, 604, 609, 611, 623, 625, 626, 627, 628,
; LOCATION: 629, 642, 643, 644, 647, 648, 651, 658, 664, 671, 673, 676,
; LOCATION: 679, 681, 691, 694, 695, 696, 698, 703, 706, 712, 713
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 714, 718
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9288

Query Match 11.8%; Score 268; DB 14; Length 721;
Best Local Similarity 89.1%; Pred. No. 4.1e-66;
Matches 294; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

QY 1350 TCATGATAGTCATCTCATCCACGAGGAGTCTTACAAATATGTATCCAGAGGCTGCA 1609
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1610 AGATAAATGSCAATTCACATTTCCACG-AFACACTGGCCAGTCAGACATCTCATAC 1668
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1669 CACATCATGATACCATCATATTTCTCCATCATCACCAACCAACCACTCTCATAC 1728
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1729 AGTCACAGCAGCGTACTCTCGGAGAGAGTGAAGATGCGGCGTGGCCCTTTGGCC 1788
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 AGTCACAGCAGCGTACTCTCGGAGAGAGTGAAGATGCGGCGTGGCCCTTTGGCC 126
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1789 TGGATGTGATATGAGTGGTGTGCTGCTGCAAAATTTGAGGATGCGGCTAGCAATTTGGTCT 1848
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1849 GCTTTTACTGAAGGCTTATCAAGTGGTTTA 1878
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 GCTTTTACTGAAGGCTTATCAAGNGTTTAA 36
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-10-198-846-1455/c
; Sequence 1455, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 1455
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2, 3, 5, 7, 9, 69, 211, 224, 360, 405, 435, 454, 488, 493
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1455
```

Query Match	11.28;	Score	253;	DB	14;	Length	504;
Best Local Similarity	87.28;	Pred.	No. 6.6e-62;				
Matches	389;	Conservative	0;	Mismatches	42;	Indels	15; Gaps
QY	1168	AGCGAAGGACCAACGAATGGAAATGAAGAAGGAGCCACTTTTCAGTCA-TCTGTCTTC	1226				
Db	497	AACCNAGCNAATGGAANAATGAAAAGAGGAGGCCACCTTTTNCAGTCAATTCTGTCTTC	438				
QY	1227	TC-AAACAATAGAAG-AAGATGCCTATTTTGATT--CCACGTGGAAGGCTCAA-CAGCT	1281				
Db	437	TCNAAACCATAGAGAAAGTGCTATTTTGTANTTTCCCAOGTGGAAGGGCTTAACCAGCT	378				
QY	1282	CTAGGAGGCGCTATTATTCAGTTTC--TTGTGGAACATGTCC-TCACATTGATCAACA	1337				
Db	377	CTAGGAGGCGCTGATTTATTTATGTTTCCCTGTTTGAACATGTCCCTCACATTGATCAACA	318				
QY	1338	ATTTAAAGATAAGAGAAGAAAAGAAATCAGAGAAGAACGTGAAAATGATCATGTGGAGAT	1397				
Db	317	ATTTAAAGATAAGAGAAGAAAAGAAATCAGAGAAGAACGTGAAAATGATCATGTGGAGAT	258				
QY	1398	TAGAAGCAGTTGTCC-AAGTATGATCTCAACTTT--CAACAATAGGAGAGAAGTAG	1453				
Db	257	TAGAAGCAGTTGTCCAAGTATGATATCTCAANCTCCACAAAANGAGAGAAGTAG	198				
QY	1454	ATACAGATGATCGAACTGAAGGCTATTATACGAGCAGACTCACAGAGCCCTCCCAGTTG	1513				
Db	197	ATACAGATGATCGAACTGAAGGCTATTATACGAGCAGACTCACAGAGCCCTCCCAGTTG	138				
QY	1514	ATTCTCAGCAGCGCTCGAGCTTGGAGAAGAGAGGTCATGATAGCTCATGCTCATCCNC	1573				
Db	137	ATTCTCAGCAGCGCTCGAGCTTGGAGAAGAGAGGTCATGATAGCTCATGCTCATCCNC	78				
QY	1574	AGGAAGTC-TACAATGAATATGTACC	1598				
Db	77	AGGAAGTCNACAATGAATATGTACC	52				

Search completed: September 12, 2003, 22:12:52
Job time : 529.999 secs

CC BCR4 gene was identified as being up-regulated in specific cancer
CC types in an oligonucleotide microarray interrogated with cDNAs
CC derived from multiple tissues. The invention provides claimed
CC methods of using BCR4 polynucleotides, polypeptides and antibodies
CC to: screen drug candidates; screen bioactive agents capable of
CC binding BCR4; evaluate the effect of a prostate cancer and/or breast
CC cancer drug; diagnose prostate cancer or breast cancer; screen for a
CC bioactive agent capable of interfering with the binding of BCR4 to
CC an antibody which binds BCR4; neutralise the effect of BCR4; treat
CC prostate cancer or breast cancer using an inhibitor of BCR4 (e.g.
CC an antibody); localise a therapeutic moiety to, or treat, prostate
CC cancer or breast cancer tissue by conjugating an antibody for BCR4
CC to a cytotoxic agent or radioisotope; inhibit prostate cancer or
CC breast cancer by administering an antisense molecule to the present
CC nucleic acid sequence; elicit an immune response by administering
CC BCR4 or a nucleic acid encoding BCR4; and determine the prognosis
CC of an individual with prostate cancer or breast cancer by
CC determining the level of BCR4 in a sample, a high level indicating
CC a poor prognosis.
XX
SQ Sequence 755 AA;

Query Match 100.0%; Score 4024; DB 23; Length 755;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARKSVLLITLFTALSVTNPLHELKAAFPQTTEKISPWNESGINVDLAISRQYHLQOL 60
DB 1 MARKSVLLITLFTALSVTNPLHELKAAFPQTTEKISPWNESGINVDLAISRQYHLQOL 60
QY 61 FRYGENNSLSVEGFRKLLQNIQIDKIKRIHHHDDHSDHEHSDHEHSDHSDH 120
DB 61 FRYGENNSLSVEGFRKLLQNIQIDKIKRIHHHDDHSDHEHSDHEHSDHSDHSDH 120
QY 121 EHSDDHSHHNAAGSKNKRKALCPDHDSDSSGKDPNRSQGGARHPHAGRNKVD 180
DB 121 EHSDDHSHHNAAGSKNKRKALCPDHDSDSSGKDPNRSQGGARHPHAGRNKVD 180
QY 181 SVSASVSTVNTVSEGTHTFTETPPGKLFPKDYSSSTPPSVTSKRSVRSLAGRT 240
DB 181 SVSASVSTVNTVSEGTHTFTETPPGKLFPKDYSSSTPPSVTSKRSVRSLAGRT 240
QY 241 NESVSEPRKGFMSNTNENPQECFNASKLLTSHGMIQVPLNATEFNLCPALINQIDA 300
DB 241 NESVSEPRKGFMSNTNENPQECFNASKLLTSHGMIQVPLNATEFNLCPALINQIDA 300
QY 301 RSLIHTSEKKAETPPKTYSLQTAWGGIAIISIFSLGVLVPLMNRVFFKLLSF 360
DB 301 RSLIHTSEKKAETPPKTYSLQTAWGGIAIISIFSLGVLVPLMNRVFFKLLSF 360
QY 361 LVALAVGTLTGDAFLHLPLPHSHASHSHSHSHEEPANEMKRGFLPFLSHLSSONTIESAYFDS 420
DB 361 LVALAVGTLTGDAFLHLPLPHSHASHSHSHSHEEPANEMKRGFLPFLSHLSSONTIESAYFDS 420
QY 421 TWKGLPALGGLYFMEVHEVTLTKQFKKKKKKKKPPENDDDVEIKKQSKYESQSLTN 480
DB 421 TWKGLPALGGLYFMEVHEVTLTKQFKKKKKKKKPPENDDDVEIKKQSKYESQSLTN 480
QY 481 EEKVDVDDRTGELRADSGQSFHSDSQPAVLEEEVEVMAHAHPQEVYNYEYPRGCKNKC 540
DB 481 EEKVDVDDRTGELRADSGQSFHSDSQPAVLEEEVEVMAHAHPQEVYNYEYPRGCKNKC 540
QY 541 HSHFHTLQSGDDLHRRHDDYHHILHHHHQNHHPHSHSQRYSREELKDAGVATLAWMI 600
DB 541 HSHFHTLQSGDDLHRRHDDYHHILHHHHQNHHPHSHSQRYSREELKDAGVATLAWMI 600
QY 601 MGDLNFDGSLGAIQAAPTEGLSSGLSTSVAVFCHPELPHGDFAVLLKAGMTVKQAVLY 660
DB 601 MGDLNFDGSLGAIQAAPTEGLSSGLSTSVAVFCHPELPHGDFAVLLKAGMTVKQAVLY 660
QY 661 NALSAMLYLGMATGIFIGHYAENVSMWIFALTAGLEMYVALVDMVPEMLHNDASDHGS 720
DB 661 NALSAMLYLGMATGIFIGHYAENVSMWIFALTAGLEMYVALVDMVPEMLHNDASDHGS 720

QY 721 RWGYFFLQAGMLLGFGLMILLISIFEHKIVPRINF 755
DB 721 RWGYFFLQAGMLLGFGLMILLISIFEHKIVPRINF 755

RESULT 3
ABR48228
ID ABR48228 standard; Protein; 755 AA.
XX
AC ABR48228;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human bladder cancer associated protein sequence SEQ ID NO:177.
XX
KW Human; bladder cancer; cytostatic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO2003003906-A2.
XX
PD 16-JAN-2003.
XX
PF 03-JUL-2002; 2002WO-US21338.
XX
PR 03-JUL-2001; 2001US-302814P.
PR 03-AUG-2001; 2001US-310099P.
PR 08-NOV-2001; 2001US-343705P.
PR 13-NOV-2001; 2001US-350666P.
PR 12-APR-2002; 2002US-372246P.
XX
PA (ECSB-) EOS BIOTECHNOLOGY INC.
XX
PI Mack DH, Aziz N;
XX
DR WPI; 2003-201532/19.
DR N-PSDB; ACC51044.
XX
PT Detecting a bladder cancer-associated transcript in a cell from a
PT patient, comprises contacting a biological sample from the patient with
PT a bladder cancer-associated polynucleotide or antibody -
XX
PS Claim 10; Page 290; 307pp; English.
XX
CC The present invention describes a method for detecting a bladder cancer-
CC associated transcript in a cell from a patient. The method comprises
CC contacting a biological sample from the patient with a polynucleotide
CC that selectively hybridises to a sequence that is 80 % identical to a
CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC encode the human bladder cancer-associated proteins given in ABR48146 to
CC ABR48242). Bladder cancer-associated sequences from the present invention
CC have cytostatic activities, and can be used in antisense gene therapy and
CC in vaccine production. The method can be used for detecting a bladder
CC cancer-associated transcript in a cell from a patient. The method is
CC useful in diagnosing or treating bladder cancer and in screening for
CC compounds that modulate bladder cancer, such as hormones or antibodies.
CC The nucleic acid molecules from the present invention may be used in
CC various screening and diagnostic methods, and for gene therapy, vaccine
CC and/or antisense/inhibition applications.
XX
SQ Sequence 755 AA;

Query Match 100.0%; Score 4024; DB 24; Length 755;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARKSVLLITLFTALSVTNPLHELKAAFPQTTEKISPWNESGINVDLAISRQYHLQOL 60
DB 1 MARKSVLLITLFTALSVTNPLHELKAAFPQTTEKISPWNESGINVDLAISRQYHLQOL 60
QY 61 FRYGENNSLSVEGFRKLLQNIQIDKIKRIHHHDDHSDHEHSDHEHSDHSDHSDH 120

Db 61 FYRGNNLSVGEFVKLLQNLGIDKIKRIHHDDHSDHEHSDHEHSDH 120
QY 121 EHHSDHSHHNNHAASGKNNKALCPDHDSDSGKDPNSOGKGAHPEHAGSRNVKD 180
Db 121 EHHSDHSHHNNHAASGKNNKALCPDHDSDSGKDPNSOGKGAHPEHAGSRNVKD 180
QY 181 SVSASEVTSTVYNTVSEGTHTFLETETPRGKLPKDPVSSSTPPSVTSKSVSLAGRKT 240
Db 181 SVSASEVTSTVYNTVSEGTHTFLETETPRGKLPKDPVSSSTPPSVTSKSVSLAGRKT 240
QY 241 NESVSEPRKGFMTSRNTNENPOECFNASKLLTSHGNGIQVPLNATEFNYLCPALINQIDA 300
Db 241 NESVSEPRKGFMTSRNTNENPOECFNASKLLTSHGNGIQVPLNATEFNYLCPALINQIDA 300
QY 301 RSLIHTSEKKAIPPKYISLQIAWVGFTAISLISFLSLGVILVPLMNVFFKLLSF 360
Db 301 RSLIHTSEKKAIPPKYISLQIAWVGFTAISLISFLSLGVILVPLMNVFFKLLSF 360
QY 361 LVALAVGTLSGDAFLHLLPHSHASHSHSHEEPAMEMKRGFLFSLSSQNIEESAYFDS 420
Db 361 LVALAVGTLSGDAFLHLLPHSHASHSHSHEEPAMEMKRGFLFSLSSQNIEESAYFDS 420
QY 421 TWKGLTALGLYPMFLVEHVLTLKQFDKDKKKNQKPPENDDDVBIKKQLSYQSLSYN 480
Db 421 TWKGLTALGLYPMFLVEHVLTLKQFDKDKKKNQKPPENDDDVBIKKQLSYQSLSYN 480
QY 481 EEKYDTRDEGYLRADSOEFSHPDSQPAVLEEEVWIAHAHPOEVNVEYVPRGCKNK 540
Db 481 EEKYDTRDEGYLRADSOEFSHPDSQPAVLEEEVWIAHAHPOEVNVEYVPRGCKNK 540
QY 541 HSHPHDTLQSDDLIHHHDYHHLHHHHQNHHPHSHSORYSREELKAGVATLAWVI 600
Db 541 HSHPHDTLQSDDLIHHHDYHHLHHHHQNHHPHSHSORYSREELKAGVATLAWVI 600
QY 601 MGDGLHNFSDGLAIGAFTGLSSGLSTSVAVFCHPELPHGDFAVLLKAGMTVKQAVLY 660
Db 601 MGDGLHNFSDGLAIGAFTGLSSGLSTSVAVFCHPELPHGDFAVLLKAGMTVKQAVLY 660
QY 661 NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFTVVALVDMVPEMLNDASDHGCS 720
Db 661 NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFTVVALVDMVPEMLNDASDHGCS 720
QY 721 RWGFFFLQAGMLLGGIMLLISIFEHKIVRINF 755
Db 721 RWGFFFLQAGMLLGGIMLLISIFEHKIVRINF 755
RESULT 4
ABU56608
ID ABU56608 standard; Protein; 755 AA.
XX
AC
XX
XX
DT 02-APR-2003 (first entry)
XX
DE Lung cancer-associated polypeptide #201.
XX
KW Lung cancer-associated polypeptide; cytostatic; emphysema;
KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW interstitial pulmonary fibrosis; asthma; bronchiectasis.
XX
OS Unidentified.
XX
XX
PN WC200286443-A2.
XX
XX
PD 31-OCT-2002.
XX
XX
PF 18-APR-2002; 2002WO-US12476.
XX
PR 18-APR-2001; 2001US-284770P.
PR 10-MAY-2001; 2001US-290492P.

PR 09-NOV-2001; 2001US-339245P.
PR 13-NOV-2001; 2001US-350666P.
PR 29-NOV-2001; 2001US-334370P.
PR 12-APR-2002; 2002US-372246P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Aziz N, Murray R;
PI
XX
XX WPI; 2003-093161/08.
DR N-PSDB; ABX76337.
XX
XX Detecting a lung cancer-associated transcript in a cell from a patient
PT for treating lung cancer, by contacting a biological sample from the
PT patient with a polynucleotide that exhibits increased or decreased
PT expression in lung cancer -
XX
XX Claim 27; Page 338; 453pp; English.
XX
XX The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridizes
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung
CC cancer-associated polynucleotides and polypeptides are used for
CC identifying a compound that modulates a lung cancer-associated
CC polypeptide, for inhibiting proliferation of a lung cancer-associated
CC cell to treat lung cancer in a patient and for treating a mammal having
CC lung cancer by administering a modulatory compound identified. The
CC methods are useful for treating lung cancer, such as small cell lung
CC cancer, non-small cell lung cancer or other benign or precancerous
CC lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
CC pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
CC and polypeptides are useful for diagnostic purposes and as targets for
CC screening for therapeutic compounds that modulate lung cancer, such as
CC antibodies. Sequences ABU56408-ABU56745 represent lung cancer-associated
CC polypeptides of the invention.
XX
SQ Sequence 755 AA;
Query Match 100.0%; Score 4024; DB 24; Length 755;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARKSVILLITLALFALSVTNPLHLEKAAAFQPTTEKISPNWESSINVDLAISTROYHLQQL 60
Db 1 MARKSVILLITLALFALSVTNPLHLEKAAAFQPTTEKISPNWESSINVDLAISTROYHLQQL 60
QY 61 FYRGNNLSVGEFVKLLQNLGIDKIKRIHHDDHSDHEHSDHEHSDH 120
Db 61 FYRGNNLSVGEFVKLLQNLGIDKIKRIHHDDHSDHEHSDHEHSDH 120
QY 121 EHHSDHSHHNNHAASGKNNKALCPDHDSDSGKDPNSOGKGAHPEHAGSRNVKD 180
Db 121 EHHSDHSHHNNHAASGKNNKALCPDHDSDSGKDPNSOGKGAHPEHAGSRNVKD 180
QY 181 SVSASEVTSTVYNTVSEGTHTFLETETPRGKLPKDPVSSSTPPSVTSKSVSLAGRKT 240
Db 181 SVSASEVTSTVYNTVSEGTHTFLETETPRGKLPKDPVSSSTPPSVTSKSVSLAGRKT 240
QY 241 NESVSEPRKGFMTSRNTNENPOECFNASKLLTSHGNGIQVPLNATEFNYLCPALINQIDA 300
Db 241 NESVSEPRKGFMTSRNTNENPOECFNASKLLTSHGNGIQVPLNATEFNYLCPALINQIDA 300
QY 301 RSLIHTSEKKAIPPKYISLQIAWVGFTAISLISFLSLGVILVPLMNVFFKLLSF 360
Db 301 RSLIHTSEKKAIPPKYISLQIAWVGFTAISLISFLSLGVILVPLMNVFFKLLSF 360
QY 361 LVALAVGTLSGDAFLHLLPHSHASHSHSHEEPAMEMKRGFLFSLSSQNIEESAYFDS 420
Db 361 LVALAVGTLSGDAFLHLLPHSHASHSHSHEEPAMEMKRGFLFSLSSQNIEESAYFDS 420

QY 421 THKGLTALGGLYFNFVHEVHLTLIKQFDKXKKKQKPPENDDDVEIKKQLSYESQLSTN 480
Db 421 THKGLTALGGLYFNFVHEVHLTLIKQFDKXKKKQKPPENDDDVEIKKQLSYESQLSTN 480
QY 481 EKVVDTDRTGYLRADSQPSHFDSQPAVLREEVNIAHAPQVYNEYVPRCKNKC 540
Db 481 EKVVDTDRTGYLRADSQPSHFDSQPAVLREEVNIAHAPQVYNEYVPRCKNKC 540
QY 541 HSHFDTLQSDDLIHHHHDYHHLHHHHQHHPHSHSQRYSGREELKDAGVATLAWVI 600
Db 541 HSHFDTLQSDDLIHHHHDYHHLHHHHQHHPHSHSQRYSGREELKDAGVATLAWVI 600
QY 601 MGDLHNPSDGLATCAFTGLSSGLSTSVAVFCHPELGHDFAVLLKAGMTVYKQAVLY 660
Db 601 MGDLHNPSDGLATCAFTGLSSGLSTSVAVFCHPELGHDFAVLLKAGMTVYKQAVLY 660
QY 661 NALSAMLAYLGMATGIGTIGHYAENVSMWIFALTAGLFWYALVDMVPEMLHNDASDHGS 720
Db 661 NALSAMLAYLGMATGIGTIGHYAENVSMWIFALTAGLFWYALVDMVPEMLHNDASDHGS 720
QY 721 RWGYFPLQAGMLLFGIMLLISIFBHKIVFRINF 755
Db 721 RWGYFPLQAGMLLFGIMLLISIFBHKIVFRINF 755
RESULT 5
ID ABG76949 standard; Protein: 755 AA.
AC ABG76949;
XX
DT 05-NOV-2002 (first entry)
XX
DE Human protein, homologous to LIV-1, designated NOV1.
XX
KW Human; SEC; NOV; immunosuppressive; hepatotropic;
KW antiinflammatory; angiogenic-associated disorder; diagnostic;
KW gene therapy; developmental disorder; immune disease;
KW signal transduction pathway disorder; metabolic disorder;
KW feeding disorder; obesity; wasting disorder; neurodegenerative disorder;
KW Alzheimer's disease; Parkinson's disease; behavioural disorder; allergy;
KW asthma; atherosclerosis; cardiomyopathy; angina pectoris;
KW autoimmune disease; retinal disease; cirrhosis; diabetes;
KW infectious disease; human immunodeficiency virus; HIV; cancer;
KW hypertension; hypotension; multiple sclerosis; urinary retention;
KW osteoporosis; Crohn's disease; ulcer; neurological disorder; anxiety;
KW haemophilia; cirrhosis; immunogen; vaccine.
XX
OS Homo sapiens.
XX
PN WO200255705-A2.
XX
PD 18-JUL-2002.
XX
PF 11-JAN-2002; 2002WO-US00609.
XX
PR 11-JAN-2001; 2001US-261013P.
PR 11-JAN-2001; 2001US-261014P.
PR 11-JAN-2001; 2001US-261018P.
PR 11-JAN-2001; 2001US-261026P.
PR 11-JAN-2001; 2001US-261029P.
PR 17-AUG-2001; 2001US-313170P.
PR 10-SEP-2001; 2001US-318410P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Mezes PS, Rastelli L, Herrmann JL, MacDougall JR, Zhong H;
PI Casman SU, Boldog F, Shinkets RA, Gorman L, Crasta OR, Mysore KK;
PI Folkerts O, Martin GB, Eisen A, Spaderna SK, Vernet CAM, Bergh C;
PI Spyttek KA, Dipippo VA, Zerhusen BD, Peyman JA, Ellerman K;
PI Stone DJ, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK;
PI Burgess CE, Edinger S;

DR WPI; 2002-590675/63.
DR N-PSDB; ABS59534.
XX
PT Human SECX/NOVX polypeptide useful for diagnosing, preventing or
PT treating disorders associated with aberrant expression or activity of
PT SECX/NOVX nucleic acids and proteins e.g., diabetes
XX
PS Claim 16; Page 81; 443pp; English.
XX
CC The invention discloses the isolated human polypeptides, and
CC polynucleotides encoding them, that have been designated SECX and NOVX.
CC The polypeptides can be used for treating, or delaying, the onset of an
CC angiogenic-associated disorder or treating a pathological state in a
CC subject, preferably a mammal. They can also be used in determining the
CC presence of, or predisposition to, a disease associated with altered
CC levels of the polypeptides and polynucleotides of any one of the 12
CC sequences (SEC1-12), for raising antibodies, for identifying an agent
CC that binds to, or that modulates the expression or activity of the
CC polypeptide, for treating or preventing a NOVX-associated disorder
CC (NOV1-8) and as a pharmaceutical composition comprising the polypeptide,
CC polynucleotide or the antibody. The polypeptides and polynucleotides are
CC useful in diagnostic applications where their amounts are assessed, or
CC for the manufacture of a medicament (e.g. gene therapy) for treating or
CC preventing disorders or syndromes such as developmental disorders, immune
CC diseases, signal transduction pathway disorders, metabolic disorders,
CC feeding disorders (including obesity), wasting disorders,
CC neurodegenerative disorders (including Alzheimer's disease and
CC Parkinson's disease), behavioural disorders, allergies, asthma,
CC atherosclerosis, cardiomyopathy, angina pectoris, autoimmune diseases,
CC retinal disease, cirrhosis, diabetes, infectious disease (bacterial,
CC fungal, protozoal and viral e.g. human immunodeficiency virus, HIV),
CC cancer (e.g. prostate cancer), hypertension, hypotension, multiple
CC sclerosis, urinary retention, osteoporosis, Crohn's disease, ulcers,
CC neurological disorders (e.g. anxiety), haemophilia or cirrhosis. They
CC may also be used as immunogens to produce antibodies specific for the
CC invention, and as vaccines. Further, they are useful for screening
CC potential agonist and antagonist compounds. The sequences presented in
CC ABG76937-ABG76956 are the human SEC1-12 and NOV1-8 proteins.
XX
SQ Sequence 755 AA;
Query Match 99.9%; Score 4020; DB 23; Length 755;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 754; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARKLSVILLITFALSVTNPLHELKAAFPQPTTEKISPWNESGINVDLAISRQYHLQOL 60
Db 1 MARKLSVILLITFALSVTNPLHELKAAFPQPTTEKISPWNESGINVDLAISRQYHLQOL 60
QY 61 FRYGENNSLSVEGFRKLQNLGIDKIKRIHHDHSDHSDHSDHSDHSDHSDHSDHSDHSDH 120
Db 61 FRYGENNSLSVEGFRKLQNLGIDKIKRIHHDHSDHSDHSDHSDHSDHSDHSDHSDHSDH 120
QY 121 EHSDDHSHHSHHNAASGKNRKALCPDHDSDSGKDPKNSQKGAHREHAGSRNVKD 180
Db 121 EHSDDHSHHSHHNAASGKNRKALCPDHDSDSGKDPKNSQKGAHREHAGSRNVKD 180
QY 181 SVSASEVTSTVNTVSEGTHTLETETPRPKLPKDVSSSTPSTSVTSKRSYRSLAGRKT 240
Db 181 SVSASEVTSTVNTVSEGTHTLETETPRPKLPKDVSSSTPSTSVTSKRSYRSLAGRKT 240
QY 241 NESVSEPRKGFMYSRNTNENPQCFNASKLLTSHGMGIQVPLNATFNFLCFAITNQIDA 300
Db 241 NESVSEPRKGFMYSRNTNENPQCFNASKLLTSHGMGIQVPLNATFNFLCFAITNQIDA 300
QY 301 RSLIHTSEKKAETPPKTYSLQIAWVGFTALSIISFLSLGLVILPLMNRVFFKLLSF 360
Db 301 RSLIHTSEKKAETPPKTYSLQIAWVGFTALSIISFLSLGLVILPLMNRVFFKLLSF 360
QY 361 LVALAVGLTSGDAFLHLLPHSHASHHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 420
Db 361 LVALAVGLTSGDAFLHLLPHSHASHHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 420

QY 421 TWKGLTALGGLYFMEFLVEHVLTLIKQPKDKKKXQKPPENDDDVEIKKQLSKTESQISTN 480
 DB 421 TWKGLTALGGLYFMEFLVEHVLTLIKQPKDKKKXQKPPENDDDVEIKKQLSKTESQISTN 480
 QY 481 EEKVDYDTRTEGYLRADSOEFSHFDSQPAVLEEEVMIHAHPQEVNVEYVPRGCKNC 540
 DB 481 EEKVDYDTRTEGYLRADSOEFSHFDSQPAVLEEEVMIHAHPQEVNVEYVPRGCKNC 540
 QY 541 HSHFHDYLGQSDLLIHHHDYHHILHHHHQNHSHSQRYSREELKDAGVATLAWMYI 600
 DB 541 HSHFHDYLGQSDLLIHHHDYHHILHHHHQNHSHSQRYSREELKDAGVATLAWMYI 600
 QY 601 MGDGLHNFSDGLAIGAATFEGLSGLSTSVAVFCHPELGHDFAVLLKAGMTVKQAVLY 660
 DB 601 MGDGLHNFSDGLAIGAATFEGLSGLSTSVAVFCHPELGHDFAVLLKAGMTVKQAVLY 660
 QY 661 NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFMTYVALVDMVPEMLHNDASDHGCS 720
 DB 661 NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFMTYVALVDMVPEMLHNDASDHGCS 720
 QY 721 RWGYFFLQAGMLLGFIMLLISIFEHKKIVRINF 755
 DB 721 RWGYFFLQAGMLLGFIMLLISIFEHKKIVRINF 755
 RESULT 6
 AA07206
 ID AAE07206 standard; Protein; 755 AA.
 XX AC AA07206;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human LIV-1-164647 protein.
 XX KW Human; LIV-1-164647; cytostatic; estrogen-inducible gene; tumour;
 KW cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;
 KW salivary gland; carcinoma; drug screening; therapy.
 XX OS Homo sapiens.
 XX PN W020015178-A2.
 XX PD 02-AUG-2001.
 XX PF 25-JAN-2001; 2001WO-US02622.
 XX PR 25-JAN-2000; 2000US-0177951.
 XX PR 10-APR-2000; 2000US-0195761.
 XX FA (GETH) GENENTECH INC.
 XX PI Goddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;
 XX WPI; 2001-502628/55.
 XX DR N-PSDB; AAD13480.
 XX PT New LIV-1 DNA and amino acids, useful for treating cancer or inhibiting
 PT the proliferation of tumor cells in mammals, e.g. breast, lung,
 PT prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma
 PT .
 XX Claim 20; Fig 2B; 150pp; English.
 XX The present sequence is human estrogen-inducible LIV-1-164647
 CC protein. LIV-1 is overexpressed in tumour tissues such as prostate,
 CC colon, lung and breast. The LIV-1 DNA and polypeptide, and the monoclonal
 CC antibody are useful for treating cancer and inhibiting the proliferation
 CC of tumour cells in mammals, e.g. breast, lung, prostate, colon, ovary,
 CC uterus, kidney, gastric or salivary carcinoma, or other tumour cell
 CC types expressing the LIV-1-164647 protein. In particular, the mammal is
 CC a human. The LIV-1 DNA and polypeptide may also be used in screening
 CC assays for drug candidates.

XX SQ Sequence 755 AA;
 Query Match 99.8%; Score 4017; DB 22; Length 755;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 754; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MARKLSVILLIFALSVNTNPLHKKAAAPQTEKISPNWESGINVDLAISRQYHLQOL 60
 DB 1 MARKLSVILLIFALSVNTNPLHKKAAAPQTEKISPNWESGINVDLAISRQYHLQOL 60
 QY 61 FYRYGNNLSVEGGRKLLQNGIDIKRIHIDHDDHSDHEHSDHEHSDHEHSDH 120
 DB 61 FYRYGNNLSVEGGRKLLQNGIDIKRIHIDHDDHSDHEHSDHEHSDHEHSDH 120
 QY 121 EHSDDHSDHSHNHAASGKNRKALCPDHDSDSSSKDPNSQSGAHRPHEASGRNVKD 180
 DB 121 EHSDDHSDHSHNHAASGKNRKALCPDHDSDSSSKDPNSQSGAHRPHEASGRNVKD 180
 QY 181 SVSASEVTSTVNTVSEGTETETPRGKLPKDYSSSTPPSVTSKSVSRLAGRKT 240
 DB 181 SVSASEVTSTVNTVSEGTETETPRGKLPKDYSSSTPPSVTSKSVSRLAGRKT 240
 QY 241 NESYSEPRKGFMYSRNTNENQECFNASKLLTSHGMIQVPLNATEFNILCPALINQIDA 300
 DB 241 NESYSEPRKGFMYSRNTNENQECFNASKLLTSHGMIQVPLNATEFNILCPALINQIDA 300
 QY 301 RSLIHTSEKKAIEPKTYSLQIAWVGGFATISILFSLGLVILVPLMNRVFFKLLSF 360
 DB 301 RSLIHTSEKKAIEPKTYSLQIAWVGGFATISILFSLGLVILVPLMNRVFFKLLSF 360
 QY 361 LVALAVGTLSGDAFLHLLPHSHASHSHSHEPAMENKRGPLFSLSSONIEESAYFDS 420
 DB 361 LVALAVGTLSGDAFLHLLPHSHASHSHSHEPAMENKRGPLFSLSSONIEESAYFDS 420
 QY 421 TWKGLTALGGLYFMEFLVEHVLTLIKQPKDKKKXQKPPENDDDVEIKKQLSKTESQISTN 480
 DB 421 TWKGLTALGGLYFMEFLVEHVLTLIKQPKDKKKXQKPPENDDDVEIKKQLSKTESQISTN 480
 QY 481 EEKVDYDTRTEGYLRADSOEFSHFDSQPAVLEEEVMIHAHPQEVNVEYVPRGCKNC 540
 DB 481 EEKVDYDTRTEGYLRADSOEFSHFDSQPAVLEEEVMIHAHPQEVNVEYVPRGCKNC 540
 QY 541 HSHFHDYLGQSDLLIHHHDYHHILHHHHQNHSHSQRYSREELKDAGVATLAWMYI 600
 DB 541 HSHFHDYLGQSDLLIHHHDYHHILHHHHQNHSHSQRYSREELKDAGVATLAWMYI 600
 QY 601 MGDGLHNFSDGLAIGAATFEGLSGLSTSVAVFCHPELGHDFAVLLKAGMTVKQAVLY 660
 DB 601 MGDGLHNFSDGLAIGAATFEGLSGLSTSVAVFCHPELGHDFAVLLKAGMTVKQAVLY 660
 QY 661 NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFMTYVALVDMVPEMLHNDASDHGCS 720
 DB 661 NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFMTYVALVDMVPEMLHNDASDHGCS 720
 QY 721 RWGYFFLQAGMLLGFIMLLISIFEHKKIVRINF 755
 DB 721 RWGYFFLQAGMLLGFIMLLISIFEHKKIVRINF 755
 RESULT 7
 ABJ05550
 ID ABJ05550 standard; Protein; 755 AA.
 XX AC ABJ05550;
 XX DT 14-NOV-2002 (first entry)
 XX DE Breast cancer-associated protein 15.
 XX KW Breast cancer; breast cancer-associated gene sequence;
 KW drug development; pharmacogenetics; biosensor development.
 XX

XX CC The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences
CC from the present invention have cytostatic activities and can be used in
CC gene therapy. The method is useful for diagnosing and treating breast
CC cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC

[illegible]

XX	RESULT 9
XX	AAU74631
ID	AAU74631 standard; Protein; 746 AA.
XX	AC
XX	AAU74631;
DT	09-APR-2002 (first entry)
XX	DE
XX	Oestrogen-regulated LIV-1 family protein U41060_Hs.
XX	LIV-1; oestrogen; cytostatic; neuroprotective; zinc homeostasis;
KW	gene therapy; apoptosis modulator; cancer; neurodegenerative disorder;
KW	apoptotic disorder; zinc-homeostasis related disorder.
XX	OS
XX	Homo sapiens.
XX	WO200196372-A2.
XX	PD
XX	20-DEC-2001.
XX	13-JUN-2001; 2001WO-GB02597.
XX	13-JUN-2000; 2000GB-0014411.
PR	14-JUN-2000; 2000GB-0014493.
PR	05-JUL-2000; 2000US-216349P.
XX	(UYCA-) UNIV COLLEGE CARDIFF.
XX	Taylor KM, Morgan HE, Nicholson RJ;
PI	WPI; 2002-106465/14.
XX	Use of a polypeptide comprising one or more consensus regions of
PT	proteins of LIV-1 family for treating disorders of zinc homeostasis
PT	such as breast cancer, neurodegenerative disorders, and for modifying
PT	apoptosis
XX	Claim 1; Figure 1; 67pp; English.
XX	The invention describes the a polypeptide comprising one or more
XX	consensus regions of proteins of LIV-1 family or its functional
CC	homologue. The polypeptide is useful in the preparation of a medicament
CC	for the treating a disease e.g. those involving disorders of zinc
CC	homeostasis, in gene therapy and for modifying apoptosis in vitro or in
CC	vivo on contact with cells. Diseases involving defects in zinc
CC	homeostasis include cancer, neurodegenerative disorders and apoptotic
CC	disorders. Recombinant proteins of the LIV-1 family (an
CC	oestrogen-regulated gene) are useful for diagnosing a zinc
CC	homeostasis-related condition in a subject which involves contacting a
CC	sample from the subject with the recombinant protein and measuring the
CC	binding of antibody to the sample. The antibody is also useful for
CC	treating a zinc homeostasis-related condition. This sequence is a
CC	member of the LIV-1 family (a gene regulated by oestrogen levels) and
CC	is useful for creating recombinant proteins for diagnosing
CC	zinc-homeostasis related conditions, described in the method of the
XX	invention.
XX	Sequence 746 AA;
XX	Sequence 746 AA;
XX	Sequence 746 AA;

QY	121	EHHSDHDDHHSHHNAASGKNNKKALCPDHDSDSGKDPNRSQKGAHPEHASGRRNVKD	181	
Db	121	EHHSD-----HNHAASGKNNKKALCPDHDSDSGKDPNRSQKGAHPEHASGRRNVKD	174	
QY	181	SVSASEVTSVYNTVSEGTHFLEIETPRCKLFPKDVSSSTPPSVTSKSRVSLAGRKT	240	
Db	175	SVSASEVTSVYNTVSEGTHFLEIETPRCKLFPKDVSSSTPPSVTSKSRVSLAGRKT	234	
QY	241	NESVSEPRKGFMTSRNTNENPOCFNASKLLTSHGMGIOVPLNATEFNYLCPALINQIDA	300	
Db	235	NESVSEPRKGFMTSRNTNENPOCFNASKLLTSHGMGIOVPLNATEFNYLCPALINQIDA	294	
QY	301	RSLIHTSEKKABIPPKTYSLQIAWVGGFIAISIIISFLSILGLVILVPLMNRVFPKFLLSF	360	
Db	295	RSLIHTSEKKABIPPKTYSLQIAWVGGFIAISIIISFLSILGLVILVPLMNRVFPKFLLSF	354	
QY	361	LYALAVGTLSGDAFLHLLPHSHASHHSHSEHPAMEMKRGPLFSLSSQNIIESAYFDS	420	
Db	355	LYALAVGTLSGDAFLHLLPHSHASHHSHSEHPAMEMKRGPLFSLSSQNIIESAYFDS	414	
QY	421	TWKGLTALGLYPMFVUEVFLTIKQFKKKKQKKPENDDDVEIKKQLSKYESQLSTN	480	
Db	415	TWKGLTALGLYPMFVUEVFLTIKQFKKKKQKKPENDDDVEIKKQLSKYESQLSTN	474	
QY	481	EKKYDTRDEGYLRADSQEPSHFDSSQPAVLEEEVMTAAHAPQEVYNETVPRGCKNK	540	
Db	475	EKKYDTRDEGYLRADSQEPSHFDSSQPAVLEEEVMTAAHAPQEVYNETVPRGCKNK	534	
QY	541	HSHPDYLGSDDLIIHHHDYHIIILHHHQNHHPHSHSQRSREELKDAGVATLAWMT	600	
Db	535	HSHPDYLGSDDLIIHHHDYHIIILHHHQNHHPHSHSQRSREELKDAGVATLAWMT	594	
QY	601	MGDLGHNFSDGLAIGAATFEGLSGLSTSVAVFCHELPHELGDFAVLLKAGMTVKQAVLY	660	
Db	595	MGDLGHNFSDGLAIGAATFEGLSGLSTSVAVFCHELPHELGDFAVLLKAGMTVKQAVLY	654	
QY	661	NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFMYVALVDMVPEMLHNDASDHGCS	720	
Db	655	NALSAMLAYLGMATGIFIGHYAENVSMWIFALTAGLFMYVALVDMVPEMLHNDASDHGCS	714	
QY	721	RWGYFFLQNAQMLLGGFMILLISIFEHKIYF	751	
Db	715	RWGYFFLQNAQMLLGGFMILLISIFEHKIYF	745	
RESULT 10				
ABJ37050				
ID	ABJ37050 standard; Protein; 752 AA.			
AC	XX			
ABJ37050;	XX			
DT	01-MAY-2003 (first entry)			
XX	XX			
DE	Human breast cancer / ovarian cancer related protein #26.			
XX	XX			
KW	Human; cytostatic; breast cancer; ovarian cancer.			
OS	Homo sapiens.			
XX	XX			
FN	WO2003000012-A2.			
XX	XX			
PD	03-JAN-2003.			
XX	XX			
PF	21-JUN-2002; 2002WO-US19773.			
XX	XX			
PR	21-JUN-2001; 2001US-300159P.			
XX	XX			
PR	27-JUN-2001; 2001US-301351P.			
XX	XX			
PA	(MILL-) MILLENNIUM PHARM INC.			
XX	XX			
PI	Veiby OP;			
XX	XX			
DR	WPI; 2003-267848/26.			

DR WPI; 1998-031740/03.
 DR N-PSDB; AAT99070, AAT99071.
 XX Assessment of metastatic risk or oestrogen responsive-ness in breast
 FT cancer - by detecting expression of pLIV1 protein
 XX
 PS Claim 1; Column 5-10; 11pp; English.
 XX
 CC This sequence is encoded by a partial sequence of the pLIV1 gene. This
 CC protein sequence is an oestrogen regulated gene associated with breast
 CC cancer. This sequence can be used in the method of the invention. The
 CC method is for determining the risk of metastasis of a female breast
 CC tumour and predicting the responsiveness of a female breast tumour to
 CC hormone treatment comprise determining if a tissue sample from the tumour
 CC expresses a polypeptide comprising at least 14 continuous amino acids of
 CC the protein encoded by this sequence. The method is useful for
 CC determining the risk of metastasis of a female breast tumour and to
 CC predict the responsive of the female breast tumour to hormone treatment.
 CC pLIV1 expression is highly predictive of tumour oestrogen receptor
 CC expression and lymph node involvement.
 XX
 XX Sequence 431 AA;
 SQ
 Query Match 54.8%; Score 2205; DB 19; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.2e-198;
 Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 328 GFAISIIISFSLGLGVILVPLMNRVFFKLLSFVLAVALVGTLSGDAPFLHLLPHSHASHH 387
 DB 1 GFAISIIISFSLGLGVILVPLMNRVFFKLLSFVLAVALVGTLSGDAPFLHLLPHSHASHH 60
 QY 388 SHSHEEPAMKRGFLPFLSHLSSONIEESAYFDSTWKGTLALGGLYFMFLVEHVLILIKOF 447
 DB 61 SHSHEEPAMKRGFLPFLSHLSSONIEESAYFDSTWKGTLALGGLYFMFLVEHVLILIKOF 120
 QY 448 KDKKKKNOKKPPENDDDVEIKKQSKYESQSTNEEKVDTDDRTGTYLRADSOEPSHFDSQ 507
 DB 121 KDKKKKNOKKPPENDDDVEIKKQSKYESQSTNEEKVDTDDRTGTYLRADSOEPSHFDSQ 180
 QY 508 QPAVLEEEVMAHAHPQEVYNEYVPRGCKNCKCHSHFDHDLTGQSDDLIHHDYHHLHH 567
 DB 181 QPAVLEEEVMAHAHPQEVYNEYVPRGCKNCKCHSHFDHDLTGQSDDLIHHDYHHLHH 240
 QY 568 HHQNHHPHSHSQSYREELKDAGVATLAWMYIMGDGLHNFSDGLAIGAAFTGSLSSGLS 627
 DB 241 HHQNHHPHSHSQSYREELKDAGVATLAWMYIMGDGLHNFSDGLAIGAAFTGSLSSGLS 300
 QY 628 TSVAVFCHLPHLGDFAVLLKAGMTVKQAVLYNALSAMLAYLGMATGIFIGHYAENVSM 687
 DB 301 TSVAVFCHLPHLGDFAVLLKAGMTVKQAVLYNALSAMLAYLGMATGIFIGHYAENVSM 360
 QY 688 WIFALTAGLFWYVALDMPVEMLHNDASDHGCSRWGFFFLONAGMLLIGRGMILLI 742
 DB 361 WIFALTAGLFWYVALDMPVEMLHNDASDHGCSRWGFFFLONAGMLLIGRGMILLI 415
 RESULT 13
 AAR98004
 ID AAR98004 standard; Protein; 431 AA.
 XX AC
 XX AAR98004;
 XX
 DT 16-OCT-1996 (first entry)
 DE Oestrogen-regulated pLIV1 gene-encoded protein.
 XX
 KW Oestrogen-regulated gene; pLIV1; breast cancer; tumour; diagnosis;
 KW marker; metastasis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 125..138

FT /label=1003
 FT /note="immunogenic peptide 1003, claim 11,
 FT page 19"
 FT Peptide 252..265
 FT /label=1004
 FT /note="immunogenic peptide 1004, claim 11
 FT page 19"
 FT Peptide 418..431
 FT /label=1105
 FT /note="immunogenic peptide 1005, claim 11
 FT page 19"
 XX CA2132500-A.
 XX 21-MAR-1996.
 XX 20-SEP-1994; 94CA-2132500.
 XX 20-SEP-1994; 94CA-2132500.
 XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
 PA
 XX Gee JM, Green CD, Manning DL, Nicholson RI;
 PI
 XX WPI; 1996-268987/28.
 DR N-PSDB; AAT33219, AAT33220.
 XX
 PT Oestrogen-regulated pLIV1 gene and deduced polypeptide - useful for
 PT predicting the propensity for metastatic spread or the
 PT responsiveness to endocrine treatment of breast tumour
 XX
 PS Claim 1; Page 14-16; 27pp; English.
 XX
 CC The polypeptide (AAR98004) encoded by a partial cDNA clone
 CC (AAT33219) corresponding to the human pLIV1 gene contains 3 peptide
 CC regions, 1003, 1004 and 1005, that were used to raise polyclonal
 CC antibodies in rabbits. Peptides 1004 and 1005 were most
 CC immunogenic, generating ELISA-positive sera at dilutions of
 CC 1/20000. Cross-reactivity studies using the 3 peptides identified
 CC 7 positive specific sera in tumour samples from 74 patients with
 CC primary breast cancer. The pLIV1 polypeptide was expressed in the
 CC cytosol of oestrogen receptor-positive epithelial cells of breast
 CC tumours. Its presence can be used to predict the propensity for
 CC metastatic spread or the responsiveness of a breast tumour to
 CC endocrine treatment. Its inactivation may be used as a means of
 CC treating breast cancer and metastasis.
 XX
 SQ Sequence 431 AA;
 Query Match 54.4%; Score 2188; DB 17; Length 431;
 Best Local Similarity 99.3%; Pred. No. 4.8e-197;
 Matches 412; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 328 GFAISIIISFSLGLGVILVPLMNRVFFKLLSFVLAVALVGTLSGDAPFLHLLPHSHASHH 387
 DB 1 GFAISIIISFSLGLGVILVPLMNRVFFKLLSFVLAVALVGTLSRDAPFLHLLPHSHASHH 60
 QY 388 SHSHEEPAMKRGFLPFLSHLSSONIEESAYFDSTWKGTLALGGLYFMFLVEHVLILIKOF 447
 DB 61 SHSHEEPAMKRGFLPFLSHLSSONIEESAYFDSTWKGTLALGGLYFMFLVEHVLILIKOF 120
 QY 448 KDKKKKNOKKPPENDDDVEIKKQSKYESQSTNEEKVDTDDRTGTYLRADSOEPSHFDSQ 507
 DB 121 KDKKKKNOKKPPENDDDVEIKKQSKYESQSTNEEKVDTDDRTGTYLRADSOEPSHFDSQ 180
 QY 508 QPAVLEEEVMAHAHPQEVYNEYVPRGCKNCKCHSHFDHDLTGQSDDLIHHDYHHLHH 567
 DB 181 QPAVLEEEVMAHAHPQEVYNEYVPRGCKNCKCHSHFDHDLTGQSDDLIHHDYHHLHH 240
 QY 568 HHQNHHPHSHSQSYREELKDAGVATLAWMYIMGDGLHNFSDGLAIGAAFTGSLSSGLS 627
 DB 241 HHQNHHPHSHSQSYREELKDAGVATLAWMYIMGDGLHNFSDGLAIGAAFTGSLSSGLS 300

QY 628 TSVAVFCHLPHELGHDFVALLKAGMTVKQAVLYNALNSAMLAYLGMATGIFIGHYAENVSM 687
 |||||
 Db 301 TSVAVFCHLPHELGHDFVALLKAGMTVKQAVLYNALNSAMLAYLGMATGIFIGHYAENVSM 360
 |||||
 QY 688 WIFALPAGLFMTVALVDMVPEMLHNDASDHGCSRWGYFFLQWAGMLLGFIMLLI 742
 |||||
 Db 361 WIFALPAGLFMTVALVDMVPEMLHNDASDHGCSRWGYFFLQWAGMLLGFIMLLI 415
 |||||

RESULT 14
 AAB56953
 ID AAB56953 standard; Protein; 397 AA.
 XX AC AAB56953;
 XX 13-MAR-2001 (first entry)
 XX Human prostate cancer antigen protein sequence SEQ ID NO:1531.
 DE Human prostate cancer; prostate cancer antigen; detection; diagnosis;
 XX Human; prostate cancer; prostate cancer antigen; immunomodulatory; muscular;
 KW neuroprotective; cytostatic; cardioactive; nephrotropic; antiinfective; gynaecological;
 KW vulnary; gastrointestinal; neural; immune; reproductive; renal;
 KW antibacterial; gene therapy; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.
 XX Homo sapiens.
 OS
 XX WO200055174-A1.
 PN 21-SEP-2000.
 PD 08-MAR-2000; 2000WO-US05988.
 PF 12-MAR-1999; 99US-0124270.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 PA Rosen CA, Ruben SM;
 XX WPI; 2000-587513/55.
 DR N-PSDB; AAF16156.
 XX Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX Claim 11; Page 1974-1975; 2338pp; English.
 XX AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.
 XX Sequence 397 AA;
 SQ Query Match 49.0%; Score 1972; DB 21; Length 397;
 Best Local Similarity 98.7%; Pred. No. 9.5e-177;
 Matches 376; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 161 SVSASEVTSTVNTVSEGHFLETETPRPGKLFPKDVSSSTPSPVTSKSRVSLAGRKT 240
 |||||

Db 1 SVSASEVTSTVNTVSEGHFLETETPRPGKLFPKDVSSSTPSPVTSKSRVSLAGRKT 60
 QY 241 NESVSPRKGFMYSNTNENPQECFNASKLLTSHGNGIQVPLNATEFNLYLCPALIINOIDA 300
 |||||
 Db 61 NESVSPRKGFMYSNTNENPQECFNASKLLTSHGNGIQVPLNATEFNLYLCPALIINOIDA 120
 |||||
 QY 301 RSLIHTSEKKAIEPPKYSIQIAWVGGFIAISISFLLGVLIVPLMNRVFFKLLSF 360
 |||||
 Db 121 RSLIHTSEKKAIEPPKYSIQIAWVGGFIAISISFLLGVLIVPLMNRVFFKLLSX 180
 |||||
 QY 361 LVLAAGTILSGDAFLHLLPHSHASHSHSHEEPAMEMKRGPLFSLHSSONIESAYFDS 420
 |||||
 Db 181 XVALAVGTLTSGDAFLHLLPHSHASHSHSHEEPAMEMKRGPLFSLHSSONIESAYFDS 240
 |||||
 QY 421 TWKGLTALGGLYFMFLVEHVTLLIKQFKDKKKKNKKPENNDDDDVEIKKOLSKYESQLSTN 480
 |||||
 Db 241 TWKGLTALGGLYFMFLVEHVTLLIKQFKDKKKKNKKPENNDDDDVEIKKOLSKYESQLSTN 300
 |||||
 QY 481 EEEKYDTRDTEGYLRADSOEPSHFDSQOPAVLEEEVMTIAHAHQEVNYEYVPRGCKNRC 540
 |||||
 Db 301 EEEKYDTRDTEGYLRADSOEPSHFDSQOPAVLEEEVMTIAHAHQEVNYEYVPRGCKKXC 360
 |||||
 QY 541 HSHFHTDLGQSDDLIHHHHDY 561
 |||||
 Db 361 HSHFHTDLGQSDDLIHHHHDF 381
 |||||

RESULT 15
 AAB87345
 ID AAB87345 standard; Protein; 831 AA.
 XX AC AAB87345;
 XX 22-MAY-2001 (first entry)
 XX Human gene 4 encoded secreted protein HDPFY41, SEQ ID NO:86.
 DE Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnary;
 KW cell culture; chemotaxis; food additive;
 KW binding partner identification.
 XX Homo sapiens.
 OS WO200118022-A1.
 PN 15-MAR-2001.
 PD 31-AUG-2000; 2000WO-US24008.
 XX 03-SEP-1999; 99US-0152315.
 PR 03-SEP-1999; 99US-0152317.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
 PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
 PI Moore PA, Shi Y, Wei Y, Florence KA;
 XX WPI; 2001-203081/20.
 DR N-PSDB; AAF91861.
 XX Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers -
 XX

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	3902	97.0	752	2	G02273	LIV-1 protein - hu
2	616	15.3	529	2	T08684	hypothetical prote
3	393.5	9.8	515	2	T23089	hypothetical prote
4	391	9.7	360	2	T19585	hypothetical prote
5	381.5	9.5	586	2	T24835	hypothetical prote
6	376	9.3	404	2	T25420	hypothetical prote
7	325	8.1	436	2	I49714	MHC H-2K/T-w5-link
8	299	7.4	346	2	S49959	probable membrane
9	271.5	6.7	433	2	T39240	hypothetical prote
10	250.5	6.2	450	2	C96704	hypothetical prote
11	208	5.2	338	2	B89623	unknown protein, 2
12	208	5.2	362	2	T19285	protein C14H10.1 [
13	194.5	4.8	735	2	T45059	hypothetical prote
14	180	4.5	3119	2	T18414	hypothetical prote
15	178	4.4	144	2	C48863	protein g377 - mal
16	171.5	4.3	668	2	A44863	trophozoite antige
17	168.5	4.2	764	2	H71607	trophozoite antige
18	168	4.2	351	1	KGZQHL	hypothetical prote
19	166.5	4.1	269	1	H71107	histidine-rich gly
20	163	4.1	65	2	D44863	gufa protein homol
21	163	4.1	283	2	C85838	trophozoite antige
22	161.5	4.0	972	2	S35521	hypothetical prote
23	160.5	4.0	477	2	T29592	DNA topoisomerase
24	158	3.9	140	2	A54523	hypothetical prote
25	157	3.9	257	2	AH1865	histidine-rich pro
26	155	3.9	279	2	H90992	hypothetical prote
27	155	3.9	295	2	T15344	hypothetical prote
28	155	3.9	330	2	D90548	hypothetical prote
29	154.5	3.8	96	2	A44971	hypothetical prote

Db 388 ASFIAGTGVIGTMTLVVHEVPEIGDFAILIQSGYSKKKAMLIQLVTA-----LGALSG 443

QY 676 IFIGHYAN-----VSMWIFALTAGLFWTVALVDMPEMLHNDASDHGSGRWGYFFL 727

Db 444 CVISLFSADADALAAASWVLPFTAGGFIYATVSVIPELLENS-----FF 492

QY 728 QNA-----GMLGFGIMLLISIFE 746

Db 493 QTVKEIFAILTGIFIMLYIAIYE 515

RESULT 4

T19585

hypothetical protein C30H6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000

C:Accession: T19585

R:Mortimore, B.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19148

A:Accession: T19585

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-360 <WIL>

A:Cross-references: EMBL:Z81044; PIDN:CAB02806.1; GSPDB:GN00022; CESP:C30H6.2

A:Experimental source: clone C30H6

C:Genetics:

A:Gene: CESP:C30H6.2

A:Map position: 4

A:Introns: 43/2; 67/3; 111/3; 149/3; 231/3

C:Superfamily: Caenorhabditis elegans hypothetical protein C30H6.2

Query Match 9.7%; Score 391; DB 2; Length 360;

Best Local Similarity 28.04; Pred. NO. 6.3e-21;

Matches 112; Conservative 58; Mismatches 136; Indels 94; Gaps 11;

QY 328 GFAISISIFSLGLVILPLMNRVFFKFLSLVALAVGLSDAFLLHP-----HSH 382

Db 15 GVISVTVSLLSLGACIGPLTKDAKHWLHFFFLAMAVSLSSDAILHIIPOVGVGDH 74

QY 383 ASHH-HSHSHHEEPAMWKRGPLFSLSSQNIIESAYFDSTW-----KGLTALG---G 430

Db 75 GHNHGSHSHHEHTDLENG-TSGHGHSHHDDSDRI--MWTVTPARKNLLRLSVLYLS 131

QY 431 LPEFELVEHLTLKQPKDKKKKQKPPENDDDVBLKKQLSKYESQLSTNEEKVDTDR 490

Db 132 IYILVFVEFFMYFRKTLHYCSPSTKSPVITSVSTHSDIKSSASNHSITGSEEDNNDS 191

QY 491 EGYLRADSOEFPFSDSQPAVLEEEVYIAHAHPQEVYNEYVPRGCKNKCHSHFDTLGQ 550

Db 192 SKERRNSVE-----LEKRRQGGH----- 211

QY 551 SDDLHHHDYHHLHHHHQNHHPHSHSQRSREBKD-----AGVATLAWVMINGDL 605

Db 212 --ELIS-----LREDGDDGTETCGKPRALILFGDGV 243

QY 606 HNFSDGLATGAFTGLSSGLSTSVAVFCHPELPHGLDFAVLLKAGMTVKQAVLYALSA 665

Db 244 HNLVDGLANGASFMTSKLGFITIAVICHPELPHGLDFAVLLSDGLSMCTAILMLISA 303

QY 666 MLAYLGMAFGIFIGHYANYSMMWIFALTAGLFMYVALVDM 705

Db 304 LTAYAGLTAIVLGR-DEEITAILAITAGMFLYVAVDM 342

RESULT 5

T24835

hypothetical protein T11F9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24835

R:Riennard, N.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19941

A:Accession: T24835

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-586 <WIL>

A:Cross-references: EMBL:Z74042; PIDN:CAA98527.1; GSPDB:GN00023; CESP:T11F9.2

A:Experimental source: clone T11F9

C:Genetics:

A:Gene: CESP:T11F9.2

A:Map position: 5

A:Introns: 86/3; 123/1; 219/2; 316/3; 538/3

Query Match 9.5%; Score 381.5; DB 2; Length 586;

Best Local Similarity 21.3%; Pred. NO. 6.1e-20;

Matches 120; Conservative 110; Mismatches 182; Indels 151; Gaps 13;

QY 286 EFNLYCPALINQIDARSCLHTSEKAEI-----PPKTY 319

Db 63 KPFYMLAALNFSISKSIEIYTIKNASQLIHQVLDHDAFLADLGNYTEAVNLDKPPAQW 122

QY 320 SQIAWVGFTAIISLISLGLVILPLMNRVFFKFLSLVALAVGLSGDAFLHLLP 379

Db 123 T-----WGIGFAIVSGCSFSAPLGLILPLCLSKSLYRIMTFLVAVGIALSGSTIFIMP 178

QY 380 HSHASHHHSHSHEEPAMWKRGPLFSLSSQNIIESAYFDSTWGLTALGGLYFMYFLVEH 439

Db 179 QAP-----HLTS-----FEHFYHKSLLIICALYAFYTVDR 210

QY 440 VLTLLKQFKDKKKKNQ-----KKPDDDDVEIKKQLSKYESQLSTNEEK 483

Db 211 MLQYILFRRRRQTKRRIHASTIASLMMNPTTKRRDNGN-----TTEETI 256

QY 484 VTDDETEGILRADSQPSH-----PDSQPAVLEEEVYIAHAHPQEVYNEYVPRG 535

Db 257 VPSEPTTILQVPLDNLHHRFRSELSDEVERTEQOEKEMALANDLEMLNNVLTAR 316

QY 536 CKNK-----CHSHFDTLGOSDGLIHH-----HDYHHILHHHHQ----- 571

Db 317 FSTRERVAVSGGLDDIEFRSPKLSHHTNGNTSFLQVINNEFRHRTMTPLSRPQSPVT 376

QY 572 -----NHPHSHSQRS-----REELKDAG--VATLAWVMINGD 603

Db 377 INVEEPKESYEMKQKSEKPGNLHODNSMSYSIRWEKVIETEPAAEVASVAMIIFGS 436

QY 604 GLHNFSDELATGAFTGLSGSLSTSVAVFCHPELPHGLDFAVLLKAGMTVKQAVLYAL 663

Db 437 SANFVDMGSMGAAPSDNLLRGLSIGIATISQOFPQELGLTALILVSGGLKTKLLFNMV 496

QY 664 SAMLAYLGMAFGIFIGHYANYSMMWIFALTAGLFMYVALVDMVPEMLHND---ASDHGS 720

Db 497 PIVLSFLGSLGVMLDSVDDSDYDEYIFAISGMYWIFLGLTIFPEIRESTNELIKENLAE 556

QY 721 RWGYFFLQAGMLLIGFGIMLLIS 743

Db 557 SILVSTILOAGILFGTTFMYFMS 579

RESULT 6

T25420

hypothetical protein T28F3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000

C:Accession: T25420

R:Mortimore, B.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z20032

A:Accession: T25420

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-404 <WIL>

A:Cross-references: EMBL:Z82285; PIDN:CAB05297.1; GSPDB:GN00022; CESP:T28F3.3

A:Experimental source: clone T28F3

C:Genetics:


```
Db 162 -----GSIHS----- 166
QY 569 HQNHHPHSQRYREELKADGAVATLAWMVGDLHNSDGLAIGAAFTBGLSSGLST 628
Db 167 -----HSHHTPQGTAB--KKAGFNMSAVLNVISGIAHHTDGLATSFYSTQVGIMT 219
QY 629 SVAVFCHPHELGHDFVALKAGMTQKQAVLYNALSAMLAVLGMATGIFIGHY----- 681
Db 220 STAVTFEIPHELGHDFAILSSGFTPPQAIRAQAVTAFGAVGTSICGWNNEIGNNSHKA 279
QY 682 ----ANVSKWIFALTAGLFIYVALVDMVPEMLHNDASDHGCSRWGYFFLONAGMLLGRGI 738
Db 280 TSSANASBELMPTTAGGLIYIATTSVVPQILHSSAPDSKLREFKKWLQLVFIQVGA 339
QY 739 MLLISIEFH 747
Db 340 MALMD--BH 346

RESULT 9
T39240
hypothetical protein SPAP8A3.03 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39240
R:Wedler, H.; Duesterboeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21742
A:Accession: T39240
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-453 <WED>
A:Cross-references: EMBL:AL117210; NID:el549906; PIDN:CAB5170.1; GSPDB:GN00066; SPDB:SH
A:Experimental source: strain 972h; clone p1 p8A3
C:Genetics:
A:Gene: SPDB:SPAP8A3.03
A:Map position: 1

Query Match 6.7%; Score 271.5; DB 2; Length 453;
Best Local Similarity 19.9%; Pred. No. 4.8e-12;
Matches 116; Conservative 94; Mismatches 191; Indels 183; Gaps 17;
QY 179 KDSVASSEVTSTVYNTVSEGTHTETPRPKLPDYSSSTPSPVTSKRVSLRAGR 238
Db 33 RDSFLSQENMKINET-----TIE-----RLF-REMTENDPSLSSKTLAELSG 77
QY 239 KNESVSPRKGFMYSRNT-----NENPQECFNASKLTHSGMGIOVPLNATEFNYLCPAI 294
Db 78 ELAKAREDLKSVLFLKNNLPVDTESSEAFTEI----- 112
QY 295 INQIDARSLHTSEKKAETPPKTYSLQIAWVGFIATSIISFLSLGVLVPLMNRVFF 354
Db 113 ----DNNSCVLNSKFEVKQSYSS--SGTNGLIATFTAPPNFIILVY----KSF 162
QY 355 KFLLSFLVALAVGTLSGDAFLHLPLPHSHASHHSHEEPAMEMKRPGLFSLHSQNIIE 414
Db 163 TSMNLNLFVAVSAGSLGDVQLPLPTVYSTN-----GGDFPASSVYSI- 205
QY 415 SAYFDSWKGTALGGLYFMFLVEHVLTILIKQFDKKKNQKPKENDDDVEIKQLSYXE 474
Db 206 -----LIGALVFFLMDKGRILIH-----RPSLSKPKKDG-----E 238
QY 475 SQLSTNE--EKVTDDR--TEGYLRADSOEPSHDSQPAVLEE--EEVMAHAHPQEVY 528
Db 239 ETSVSNKFSASSTQDVKGVEGLRKNRKNVDQNSKGHEFDLIRHVEEV-----SEY 291
QY 529 NEVYPRGCKNKHSHFHTDLQSDDLIHHDHYHHLHHHHQNHHPHSQRSREELK 588
Db 292 ND----- 293
QY 589 DAGVATLAWMVGDLHNSDGLAIGAAFTBGLSSGLSTSVAVFCHPHELGHDFAYILL 648
Db 294 ----KTVYVLNLLCDSFNFMDGLAITSFAFTNTSIGISTTFAVLLHETPABIGDLAIL 349
```

```
QY 649 KAGMTVQAVLYNALSAMLAVLGNATGIFI-----CHYAENVSMWIFALTAGLEM 698
Db 350 RNYTKRSQVILQMTITWYTGLLGAIIVATYITASSSSPYGSLQLQLEDKLLPFTAGGFL 409
QY 699 YVALVDMVPEMLHNDASDHGCSRWGYFFLQONAGMLLGFIMLLI 742
Db 410 XTALGVFPPELLEINLSKGLGNWITYALYMMFIVGGFSFLYYV 453

RESULT 10
C96704
unknown protein, 23065-20358 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96704
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alc
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; NUID:21016719; PMID:11130712
A:Accession: C96704
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-450 <STO>
A:Cross-references: GB:AE005173; NID:g6553886; PIDN:AAF16552.1; GSPDB:GN00141
C:Genetics:
A:Gene: T23K23.5
A:Map position: 1
```

```
Query Match 6.2%; Score 250.5; DB 2; Length 450;
Best Local Similarity 18.2%; Pred. No. 1.6e-10;
Matches 124; Conservative 91; Mismatches 164; Indels 303; Gaps 22;
QY 101 DBEHSDDHRSDDHSHSDHDDHSHHNAASGNKRKALCPDHDSDSSGKDPKN 160
Db 36 DHVHH--HGGGCGSHS--DHDHDDHDDH----- 60
QY 161 SQCKAGRPHASGRNVKDSVASSEVTSTVYNTVSEGTHTETPRPKLPKPDWSS 220
Db 61 -----VKKTTAKVEM----- 70
QY 221 STPPSVTSKRSVRLAGKNTNVSVPKGFMYSRNTNENPQECFNASKLTHSGMGIOV 280
Db 71 -----KLPEELAE-----BEDMRLC-----GFG---- 88
QY 281 PLNATEFNYLCPAIQIDARSLHTSEKKAETPPKTYSLQIAWVGFIATSIISFLSL 340
Db 89 -----PCL-HDHDHSSSTLTGFL--WLNALGCSLLVLSL 123
QY 341 LGVILVPLMNR---VPEKFLLSFLVALAVGTLSGDAFLHLPLPHSHASHHSHEEPAME 397
Db 124 IGVLLPIMFALLILVLEKMLSFVYSCAI-----MQELCWE 159
QY 398 MKRGPLFSLSSQNIIEESAVFDSTWGLTALGGLYFMFLVEHVLTLI----- 444
Db 160 ML---PFTNCPMLLVATILMTMTMTIMILIRLHHTHILYKLCCLDCLFLGLWS 216
QY 445 -----KQFKKKKKKKPKPENDD--VEIKQLSKYESQLSTNEEKVTDTRTEG 492
Db 217 SFLWSWCSSKKIKDEGHDHNLQSSSDAIVNSSEKYSGGSTDKSLRKKTKSASDATD- 275
QY 493 YLRADS--QEPSHFDSDQPAVLEEEVMAHAHPQEVYNEVYPRGCKNKHSHFHTLQ 550
Db 276 --KDSGTEITSDGSKDKPEQVET-----R 298
QY 551 SDDLIIHHHDYHHLHHHHQNHHPHSQRSREELKADGAVATLAWMVGDLHNSD 610
```


Result No.	Query			ID	Description
	Score	Match	Length		
1	455	11.3	449	CSUP_DROME	O9v3a4 drosophila
2	393.5	9.8	515	KEAL_CAEEL	O9xt07 caenorhabdi
3	389	9.7	476	KE4_MOUSE	Q31125 mus muscullu
4	376	9.3	404	IGKX_CAEEL	O9xuc4 caenorhabdi
5	355.5	8.8	469	KE4_HUMAN	O92504 homo sapien
6	299	7.4	346	YIC3_YEAST	P40544 saccharomyc
7	282.5	7.0	352	KE4_BRARE	O9pub8 brachydanio
8	273	6.8	355	HRP6_DROME	O9vaf0 drosophila
9	168	4.2	351	HRPX_PLAIO	P04929 plasmodium
10	165.5	4.1	338	IARL_ARATH	O9m647 arabidopsis
11	161.5	4.0	972	TOPL_DROME	P30189 drosophila
12	157	3.9	302	HYPE_BRAJA	O45257 bradyrhizob
13	154	3.8	337	ZNUA_HAETN	P44526 haemophilus
14	147.5	3.7	291	ZUPT_CAMJE	O9pin2 campylobact
15	146	3.6	549	DSX_DROME	P23023 drosophila
16	145	3.6	254	GUPA_MYXXA	O05916 myxococcus
17	142	3.5	1709	CHDL_HUMAN	O16466 homo sapien
18	141	3.5	950	URBL_USTMA	P40349 ustilago ma
19	140	3.5	332	HRP1_PLATFA	P05227 plasmodium
20	140	3.5	732	YLH3_SCHPO	O9hgq3 schizosacch
21	139.5	3.5	2026	CYAA_YEAST	P08678 saccharomyc
22	137.5	3.4	1711	CHDL_MOUSE	P40201 mus muscullu
23	135.5	3.4	149	EGGS_SCHMA	P08016 schistosoma
24	135	3.4	722	HNW2_DROME	P22808 drosophila
25	134	3.3	503	YKR5_YEAST	P34240 saccharomyc
26	134	3.3	1093	PI4K_DICTDI	P54677 dictyosteli
27	133	3.3	657	KNOM_PLAFN	P06719 plasmodium
28	131.5	3.3	1402	SALM_DROVI	P39806 drosophila
29	130.5	3.2	257	ZUPT_ECOLI	P24198 escherichia
30	129.5	3.2	626	DALI_DROME	Q24114 drosophila
31	129	3.2	257	ZUPT_SALTY	O8xgr4 salmonella
32	129	3.2	285	ZUPT_CLOPE	O8xm98 clostridium
33	127.5	3.2	390	PC_DROME	P26017 drosophila

"A putative transmembrane protein with histidine-rich charge clusters encoded in the H-2K/tw5 region of mice."; Mol. Cell. Biol. 10:138-145(1990). [2]

SEQUENCE FROM N.A.
STRAIN=129/SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY IN EMBRYONIC CARCINOMA
CC CELLS, BUT WEAKLY IN ADULT TISSUES.
CC -!- SIMILARITY: BELONGS TO THE K64/CATSUP FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; M32010; AAA37767.1; .
DR EMBL; AF100956; AAC69903.1; .
DR GMD; MG1-95909; H2-K64.
DR InterPro; IPR003689; Zn_transpt_Zip.
DR Pfam; PF02535; Zip; 1.
DR Transmembrane; Glycoprotein.
KW TRANSMEM 7 27
FT TRANSMEM 146 166 POTENTIAL,
FT TRANSMEM 177 197 POTENTIAL,
FT TRANSMEM 222 242 POTENTIAL,
FT TRANSMEM 393 413 POTENTIAL,
FT TRANSMEM 417 437 POTENTIAL,
FT DOMAIN 27 136 HIS-RICH.
FT DOMAIN 252 274 HIS-RICH.
FT CARBOHYD 337 337 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 367 367 V -> L (IN REF. 1);
FT CONFLICT 403 476 ACALLTEGGANDSDVAGGAGPGWLPPTAGGTGYVAVSVL
PELLREASPLQSLLEVLGLLGGMVMVLAHLE -> RVFPL
SPREGQWTWVGTVGVLAGSCHSIQADLST (IN REF.
1).
SEQUENCE 476 AA; 50656 MW; 48214438BE44919B CRC64;

Query Match 9.7%; Score 389; DB 1; Length 476;
Best Local Similarity 23.0%; Pred.No. 7.4e-20;
Matches 156; Conservative 66; Mismatches 175; Indels 282; Gaps 26;

QY 93 HPDHHDHSDE--HHSDHERSD-----HEHSHDSHSDHDDHSH--HNHAASGNKRKA 145
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 55 HHGSHSGHSHEDFHGHCHTHESIWGHASHSDHGDHNSPEELHHGHSGHS----- 104
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 146 CPDHDSSGDKDPNRSOGK-AHRPHASGRNRKNVDKSYSASEVTSTVYNVTSSEGFLET 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 105 ---HDSLHG-----GHGHAHR-EHSHG----- 123
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 205 IETPRCKFLPKDVSSPTPSVKSRVSRLAGRTKNESVEPRKGFMYSRNTNENPOEC 264
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 124 -----TSREAG----- 129
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 265 FNASKLLTSHGNGIOVPLNATENYLCPALINDARSCLHTSEKKAEIPKYSLQIA 324
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 130 -----APGIKHLD-----TTWLWAY 145
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 325 WVGCGTAISITFSLLGLVILVPL-MNRVFKEFLSFVALAVGTLGDAFLHLPHS-- 381
:||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 146 ALGATVLSAAPFFVL---FLIPIVSNRPRLSLLOILLFSAAGGLGDALFIHLIPALE 202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 382 -HASHH-----HSHSHEEPAMKRGPLFSLHSQNIIEFSAYFDSTWKGLTAGGLGYFM 434
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 203 PH-SHAPEOPCHGSHS-----GGPTIS-----VGIWVIGIVAF 238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :

Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.

[4]

RL SEQUENCE FROM N.A.
RP TISSUE=SKIN;
RC MEDLINE=2238857; PubMed=12477932;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.L., Collins P.S., Wagner L., Scheamen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.C., Rubin G.M., Hong L.,
RA Stapleton M.J., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Tohyuki S., Carninci P., Prange C.,
RA Raba S.S., Locquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Posak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.D., Hulyk S.W.,
RA Villalón D.K., Muny D.N., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Butterfield A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Rodriguez Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.; et al.
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: MAJOR EXPRESSION IN PLACENTA, LUNG, KIDNEY
CC AND PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE KB4/CATSUP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenselib@sib.ch).
CC -----
DR EMBL; D82060; BAAL1528.1; +
DR EMBL; AF117221; AAD12305.1; -
DR EMBL; AL031228; CAA20238.1; -
DR EMBL; BC00645; AAH00645.1; -
DR Genew; HGNC:4927; HK4.
DR MIM; 601416; --
DR GO; GO:0005624; C:membrane fraction; TAS.
DR InterPro; IPR003689; zn_transpt_zip.
DR Pfam; PF02535; zip; 1.
KW Transmembrane; Glycoprotein.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 138 138 POTENTIAL.
FT TRANSMEM 169 169 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 417 436 POTENTIAL.
FT DOMAIN 30 114 HIS-RICH.
FT DOMAIN 238 263 HIS-RICH.
FT CARBOHYD 330 330 N-LINKED (GLCNAC. ..) (POTENTIAL).
FT CONFLICT 7 7 A -> G (IN REF. 1 AND 2).
FT CONFLICT 280 280 E -> G (IN REF. 1 AND 2).
FT CONFLICT 376 376 S -> T (IN REF. 1 AND 2).
FT CONFLICT 397 469 CALLTEGAGVSRITAGAGPGWLPPTFGGFIYKATVSFLP
FT ELLEAPSLQSLLEVLIGLVGMVLMVLIABLE -> VPFSL
FT KEEOATVKLVQVGVLAGSCHLLQVALST (IN REF. 1
FT AND 2).
SQ SEQUENCE 469 AA; 50118 MW; 6504A1EF5AA6A5B9 CRC64;

Query Match 8.8%; Score 355.5; DB 1; Length 469;
Best Local Similarity 20.6%; Pred. No. 1.6e-17;
Matches 145; Conservative 68; Mismatches 159; Indels 333; Gaps 23;

QY 91 HTRHDDHDSDHEHSHDSRHS-D-HEHHSDEHT-HSDDHSHHNHAASGKNKKRALCPD 148
Db 49 HSHDFHHGHSHARGHGTHIISIVGHGTHDHDHGHSHDLHHGHSG----- 97

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 245881; CRA6969.1; .
 DR PIR; S49959; S49959.
 DR SGD; S0001285; YII023C.
 DR InterPro; IPR003689; zn_transpt.zip.
 DR Pfam; PF02535; Zip: 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 100 120 POTENTIAL.
 FT TRANSMEM 127 147 POTENTIAL.
 FT TRANSMEM 203 223 POTENTIAL.
 FT TRANSMEM 253 273 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 346 AA; 37418 MW; 645960E194778006 CRC64;
 Query Match 7.4%; Score 299; DB 1; Length 346;
 Best Local Similarity 23.1%; Pred. No. 9.9e-14; Indels 164; Gaps 10;
 Matches 99; Conservative 51; Mismatches 115;
 QY 330 IAIISILFSLGVLVLP-LMNRVFFKLLSFLVALAVGTLSGDAFLHLLPHSHASHHSHS 388
 Db 71 VAILIQLMPLFVLPVGRKNDRASITLSLVFSGLTGLDILLHVIPES----- 123
 QY 389 HSHEEPAMKRGFLSHLSQNTESAYFDSTWKGTLGALGLYPMFLVERVLTLLKQPK 448
 Db 124 -----LSGVT-----DVTWVGGAIFLGIS-FTLT----- 147
 QY 449 DKKKNQKPKENDDDVEIKQLSKYESQLSTNEEKVTDTRTEGYLRADSQEPHFSDSQ 508
 Db 148 DKTWRIISGTSNDD----- 161
 QY 509 PAYLEEEVMIHAHQVEYNEYVPRCKNKHSHPHDTLGOSDDLIHHHEDYHHLHH 568
 Db 162 -----GSIHS----- 166
 QY 569 HQNHHPHSHSORYSREELKDAGVATLAWVIMGDLGNFSDGLAIGAAFTGGLSLST 628
 Db 167 -----HSRSHTPQOTAE--KRAGFNMSAYLNVISGTAHHTDGLATSFYSSTQVGIMT 219
 QY 629 SVAYFCHPELHFGDPAVLKAGMTVKQAVLYNALSMAYLGMATGIFIGHY----- 681
 Db 220 SIATVTFEIPHELGDFAILLSGFTFPQAIRAQAVTAFAVGVTSIGQWNEIGNNSHKA 279
 QY 682 ---AENYSMIFALTAGLFMVALVDVPEMLNDASDHGCSRWGYFFLQAGMLLGF 738
 Db 280 TSSANASAMLPFTTAGGLIYIATSVYVQILHSSAPDSKLREFFKKWALQLVFVGF 339
 QY 739 MLLSIFEH 747
 Db 340 MALMD--EH 346
 RESULT 7
 ID KE4_BRARE STANDARD; PRT; 352 AA.
 AC Q2PUB8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Histidine-rich membrane protein ke4 homolog (Fragment).
 GN HKE4 OR KE4.

OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Murray B.W., Snelmann H., Klein J.;
 RT "Identification of a homolog of the human HKE4 gene in zebrafish."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE KE4/CATSUP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF196345; AAP05821.1; .
 DR ZFIN; ZDB-GENE-991110-20; Ke4.
 DR InterPro; IPR003689; zn_transpt.zip.
 DR Pfam; PF02535; Zip: 1.
 KW Transmembrane; Glycoprotein.
 FT TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 215 235 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT DOMAIN 24 105 HIS-RICH.
 FT DOMAIN 177 217 HIS-RICH.
 FT CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 352 352
 SQ SEQUENCE 352 AA; 37922 MW; C8C8C60F6D2BA8A6 CRC64;
 Query Match 7.0%; Score 282.5; DB 1; Length 352;
 Best Local Similarity 19.2%; Pred. No. 1.4e-12;
 Matches 118; Conservative 43; Mismatches 113; Indels 339; Gaps 17;
 QY 91 HIHHDH-----HSDDEHH-SDEHSHSDEHSDH 120
 Db 24 HSHHHGGGCGCHGSHGAKMHGASKWAPANLPHAEHEHVDHGHHDHGH 83
 QY 121 EH-HSD-HSH-HSHAASGKNRKALCPDHDSDGKDPKNSQKGAHPEHAGSR 176
 Db 84 AHSGLDHDGHAKKHGA-----HDGAESKKYVE-----AGKR 119
 QY 177 NKDSVSASEVTSTVYNTVSGTHLETIETPRPKLPKDVSSSTPSPVTSKRSRLA 236
 Db 120 MWVE----- 123
 QY 237 GRKTNSVSEPRKGFMYSRNTNENPQCFNASKLLTSHGMGIQVPLNATEFYLCFAIN 296
 Db 124 -----LWQAIGATLLISAAPFLIL---FLIPVQSNDDQ 156
 QY 356 FLLSFLVALAVGTLGSDAFLLHLLPHSHASHHSHSHEEPAMKRGFLPHLSQNIIES 415
 Db 157 NLKVLISFASGGLGDAFLHLLPHALEPHSH---HSQPHSEESHGSHGSHGSHA 213
 QY 416 AYFDSWKGTALGGLYFMFLVERVLTLLK-----QFKDK-----KKKNQK 457
 Db 214 AHCHMSVGLVWLGGIVAFVLYVEKRVLLAGCHSHSHSAPAKSKDSDEEDKKGQK 273
 QY 458 PENDDDVEIKQLSKYESQLSTNEEKVDT--DDRTGYLRADSQEPHFSDSQPVALEEE 515
 Db 274 GEKD-----KVSQKPTKTKVTETSSDIKVSGLN----- 303

QY 516 EYMTAHAPQEVYNEYVPRGCKNCHSHFDITLQSDLLHHHHYHHHLLHHHHQNHHP 575
 Db 304 ----- 303
 QY 576 HSHSQYSREELKAGVATLAWVINGDLNFDGLAIGAAFTGLSSGLSTSVAVTCH 635
 Db 304 -----LAADFTHNFTDGLAIGASFLVGAFAVTTITILL 339
 QY 636 ELPHEIGDFAVLL 648
 Db 340 EWPHEIGDFAVLL 352
 RESULT 8
 ID Y816 DROME STANDARD; PRT; 355 AA.
 AC QVAF0:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CG7816.
 GN CG7816.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson X., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegvam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE KE4/CATSUP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003771; AN14195.1; -.
 DR FlyBase; FBgn0039714; CG7816.
 DR InterPro; IPR003689; zn_transp_zip.
 DR Pfam; PF02535; zip; 1.
 KW Hypothetical protein; Transmembrane; Glycoprotein.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 79 99 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 301 321 POTENTIAL.
 FT CARBOHYD 4 4 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 355 AA; 38870 MW; 01527C0390741FE8 CRC64;
 SQ SEQUENCE 355 AA; 6.8%; Score 273; DB 1; Length 355;
 Query Match Best Local Similarity 20.8%; Pred. No. 6.8e-12;
 Matches 96; Conservative 76; Mismatches 162; Indels 128; Gaps 14;
 QY 299 DARSCLHTSEKAEIPPKYISLQIA-WVGFTAIISIIISLGLVILVPL---MNRVFF 354
 Db 9 DEHIAIYISNMDQYMPFKSYFEYTPWFSLLGSVIGLSGIFPLIIITPEEKMAKEG 68
 QY 355 K-----FLLSFLVALAVGTLSDAPFLHLLPHSHASHSHSHEEPAMEMKRGPLFSHLS 409
 Db 69 KDPADSKLLRVLSFAVGGLGLDVLHLLPAWEG-----DNQDPS-----SHPSL 114
 QY 410 QNIEESAYFSDTWKGLTGGLYFMFLVHEVHLILKOFKKKKKKPKPENDDDVEIKQ 469
 Db 115 RS-----GLWLSGLIIFTVEIKFS----- 135
 QY 470 LSKVESQLSTNEKVTDDRTGILRADSQPSHFDSDQPAVLEEEYVIAHAHPQEVYN 529
 Db 136 -----GYASADENP-----QPKCVETANCLL-RRGGGLPE 166
 QY 530 EYVPRGCKNCHSHFDITLQSDLLHHHHYHHHLLHHHHQNHHPHSHSQYSREELK 589
 Db 167 GETSESGGACDI---EDVGKVCFL-----REQEOKSKERKE 200
 QY 590 AGVATLAWVINGDLNFDGLAIGAAFTGLSSGLSTSVAVTCHPHEIGDFAVLLK 649
 Db 201 QPKKVAGYVLLNLSIDNFTHGLAVAGSFLVSPFRHGLATFATFALLHPEHVGDFAILLR 260
 QY 650 AGMT---VKQAVLYNALSAMLAYLGMATGIFTIGHYAEVNSMMIFALTAGLFMYVALYDMV 706
 Db 261 SGFSNDAAARQLITAGAGLLGALVAIGSGVTSMEARTSWIMPTTAGGFLHALVTVL 320
 QY 707 PEMLNHDSHGCSRMWGFFLQNAQMLGFGIMLLIS-IFEH 747
 Db 321 PDLLEKEERKESTK-----QLLALVFGIALMAVMTLFEH 355
 RESULT 9
 ID HRPX_PLALO STANDARD; PRT; 351 AA.
 AC P04929;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DE Histidine-rich glycoprotein precursor.
 OS Plasmodium lophurae.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5853;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85061618; PubMed=6095114;
 RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.;


```

DR EMBL; AC012563; AAG52008.1; ALT_SEQ.
DR EMBL; AY072108; AA159930.1; -.
DR InterPro; IPR003689; zn_transpt.zip.
DR Pfam; PF02535; zip; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT DOMAIN 37 60 HIS-RICH.
FT DOMAIN 161 242 HIS-RICH.
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 236 242 POLY-HIS.
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 338 AA; 38954 MW; 07095C8B5BC01DAB CRC64;

Query Match
Best Local Similarity 18.7%; Pred. No. 0.00022;
Matches 85; Conservative 49; Mismatches 104; Indels 217; Gaps 20;

QY 101 DHEHSDHERHSDHEHSDHEHSDHEHSDHSHHNAAGKKNKALCPDHDSDSGKDPN 160
Db 36 DHVHH--HGCGSHSH--DHDHDDHDDH----- 60
QY 161 SQKGARHPEHAGSRNRNVDVSASEVTSTVNTVSEGTHTLETITPREGKLPKDVSS 220
Db 61 -----VKTTAKVEM----- 70
QY 221 STPPSVTSKSRVSLAGRTNSVSEPRKGFMYSRNINENPQCFNASKLITSHGMGIQV 280
Db 71 -----KLPEELAE-----EEDMLC-----GFG--- 88
QY 281 PLNATENYLCPALINQIDARSLIHTSEKAEIPPKYTSLOIATWVGFGFIALSIISPLSL 340
Db 89 -----PCL-HDDHSSSTLGFAL---WLNLGCSLIVSLASL 123
QY 341 LGVILVPLMNRVFF-----KFLSLFVALAVGTLSGDAFLHLPHS-----HAS 384
Db 124 ICLVLLPIM---FVQGPKSQWFDLSALFGAGMLGDAFLHLQHPHAFGGGSHSHNDHEN 180
QY 385 H-HSHSHSEEPAMKRGPLFHLSSQNIIESAYFTDWTGLTALGGLYFMFLVHVLTL 443
Db 181 HDHDDHSD-----SPSHS-SIQDLS-----VGLSVLAGIVVFLVLEKLVRY 223
QY 444 I-----KQFKDKKKKKKKPPENDDD--VEIKKOLSKYSQSLSTN 480
Db 224 VEENSSGNTWGHSHHHHAGSKKLKDEGDNNDLQSSSDALVNSSEKVSQGGTDXSLR 283
QY 481 EEKVDTRTEGYLRADS--QPPSHFDSQQPAVLE 513
Db 284 KRKTSASDATD--KSDSGTEITSQKSKDKEQVE 315

RESULT 11
ID TOPI_DROME STANDARD; PRT; 972 AA.
AC P30189; Q9YXW6;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2).
GN TOPI OR CG6146.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117086; PubMed=1335568;
RA Hsieh T.-S., Brown S.D., Huang P., Postel J.;
RT "Isolation and characterization of a gene encoding DNA topoisomerase
  I in Drosophila melanogaster.";

```

```

RL Nucleic Acids Res. 20:6177-6182(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=96354910; PubMed=8769417;
RA Zhang C.X., Lee M.P., Chen A.D., Brown S.D., Hsieh T.-S.;
RT "Isolation and characterization of a Drosophila gene essential for
  early embryonic development and formation of cortical cleavage
  furrows.";
RL J. Cell Biol. 134:923-934(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Anatolides P.G., Scherer S.E., Holt R.A., Evans C.A., Gocayne J.D.,
  Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
  Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
  Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
  Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
  de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
  Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
  Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
  Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
  Klamel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
  Szivaskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
  Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000)
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----

```

DT	15-SEP-2003 (Rel. 42, Last annotation update)
DE	Hydrogenase nickel incorporation protein hypB.
DN	HYPB OR BLI6931.
OS	Bradyrhizobium japonicum.
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CC	Bradyrhizobiaceae; Bradyrhizobium.
OX	NCBI_TaxID=375;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-USDA 110;
RX	MEDLINE=94137733; PubMed=8305450;
RA	Fu C., Maier R.J.;
RT	"Nucleotide sequences of two hydrogenase-related genes (hypA and
RT	hypB) from bradyrhizobium japonicum, one of which (hypB) encodes an
RT	extremely histidine-rich region and guanine nucleotide-binding
RT	domains."
RL	Biochim. Biophys. Acta 1184:135-138(1994).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-USDA 110;
RX	MEDLINE=22484998; PubMed=12597275;
RA	Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA	Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA	Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA	Tabata S.;
RT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT	Bradyrhizobium japonicum USDA110.";
RL	DNA Res. 9:189-197(2002).
CC	-!- FUNCTION: MAY WORK IN THE MOBILIZATION OF NICKEL INTO HYDROGENASE
CC	ENZYME. BINDS 9 NICKEL IONS PER MOLECULE.
CC	-!- SIMILARITY: BELONGS TO THE HYPE/HUPM FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; L24513; AAA17763.1; -.
DR	EMBL; AP005960; BAC52196.1; -.
DR	InterPro; IPR004392; HYPB.
DR	InterPro; IPR002894; HYPB_Ureg.
DR	Pfam; PF01495; HYPB_Ureg_1.
DR	TIGRFAMs; TIGR00073; hypB; 1.
KW	Metal-binding; Nickel; Complete proteome.
FT	DOMAIN 16 54 HIS-RICH.
FT	CONFLICT 72 72 A -> T (IN REF. 1).
FT	SEQUENCE 302 AA; 32708 MW; D3B5F5424AB9AA CRC64;
QY	Query Match 3.9%; Score 157; DB 1; Length 302;
Db	Best Local Similarity 21.4%; Pred.No. 0.00074;
Db	Matches 67; Conservative 27; Mismatches 101; Indels 118; Gaps 11;
QY	92 IHHDHSHSDHEHHSDHERHSDHSHDSHDSHSHHNHAAS----- 137
Db	
Db	14 IEHADHDDHGHDH----GHGHHHHHHGHGDQHHDHDAHGDAGLLDCGANPAQG 69
QY	138 -----GNKRRKALCPDH-----DSDSSGGKDPRNSQKG 164
Db	
Db	70 KIAGMSDRIIQVRDILGKNDR-LAADNRARFRADENVLAFLNVSPGAGKTSLLVRAV 127
QY	165 GAHRPEHSG-----RRNVKDVSNASEVTSTVNTVSECTHF-----LETIEIPR 209
Db	: : : : : : :
Db	128 SELKDFALGVIEGQQQTSDAEIRATGPVAIQNT-GKGCHLDRAAMVEAYDKLPWLN 186
QY	210 PGKLFPKDYSSSTPPSVTSKVSXSLGRKTNESYSEPKGFMTSRNTNENPQECFNASK 269
Db	: : : : : :
Db	197 GGLLFITENVNLVPAEDLGGEAKIVVFSTFEGEDKKPLK-----YPDMPFASS 235
QY	270 LLASHGMGIQVPLNATEFYNLPCALINQIDAASC-----LIETSEKKAETPPKTYLSQI-- 323
Db	: : : : : :

CC	EMBL; U32698; AAC21794.1; -	
DR	PIR; D64049; D64049.	
DR	TIGR; H10119; -	
DR	InterPro; IPR006127; SBP_bac_9.	
DR	Pfam; PF01297; SBP_bac_9; 1.	
KW	Transport; Periplasmic; zinc transport; zinc; Metal-binding; Signal;	
KW	Complete proteome.	
FT	SIGNAL	1 23
FT	CHAIN	24 337
FT		HIGH-AFFINITY ZINC UPTAKE SYSTEM PROTEIN
FT		ZNUA.
FT	DOMAIN	115 163
FT	METAL	57 57
FT	METAL	171 171
FT	METAL	235 235
FT	SEQUENCE	337 AA; 3DB845AB8F06FCB CRC64;

KW Transport; Zinc transport; Transmembrane; Complete proteome.

SQ SEQUENCE 291 AA; 31462 MW; 35A0E51E408E1CF2 CRC64;

Query Match 3.7%; Score 147.5; DB 1; Length 291;
 Best Local Similarity 25.4%; Pred. No. 0.0033;
 Matches 44; Conservative 32; Mismatches 74; Indels 23; Gaps 4;

QY 574 HPHSQRYSRRELKAGVATLAWYIMGDLNFSGDGAIAGAAFTGLSSGLSTSVAVF 633
 || : : || : : : : || : : : : || : : : : || : : : : || : : : : ||
 130 HPGEKLHQINTKALKRTGIFT-----ALAIAHNPEGFATFISSLNLFGLAIAIAVA 184

QY 634 CHELPELGDFAVLKAGMTVKQAVLYNALSLAMLAYLGNATGI-----FCHGAENVSMW 588
 || : || : || : || : || : || : || : || : || : || : || : || : || : ||
 185 IHNIPGLAVSPYIHATDKKKAFTYSALSFGFAELGAFVGLALILFFDITLTAIS-- 242

QY 689 IFALTAGLPMYVALVDWVPEMLHNDASDHGCSRWGYFFLQNGMLGFGIMLL 741
 || : || : || : || : || : || : || : || : || : || : || : || : || : ||
 243 -FAVTAGIMVFISLDELIPRAKYDKHDSL-----YGLIAGNATMAL 284

RESULT 15

DSX_DROME STANDARD; PRT; 549 AA.

AC P23023; P23022; Q9VHV0;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Doublesex protein.
 GN DSX OR CG11094.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS FEMALE AND MALE).
 RC TISSUE=Larva, and Pupae;
 RX MEDLINE=89168451; PubMed=2493994;
 RA Burtis K.C., Baker B.S.;
 RT "Drosophila doublesex gene controls somatic sexual differentiation by
 RT producing alternatively spliced mRNAs encoding related sex-specific
 RT polypeptides."
 RL Cell 56:997-1010(1989).
 RP [2]

SEQUENCE FROM N.A. (ISOFORM MALE).

RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA Cherry J.M., Cusum B.A., Butler H., Cadenat E., Center A., Chandra I.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
 RA Palazzolo R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassaman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP DNA-BINDING.
 RX MEDLINE=91330881; PubMed=1907913;
 RA Burtis K.C., Coschigano K.T., Baker B.S., Wensink P.C.;
 RT "The doublesex proteins of Drosophila melanogaster bind directly to a
 RT sex-specific yolk protein gene enhancer.";
 RL EMBO J. 10:2577-2582(1991).
 RN [4]
 RP DNA-BINDING DOMAIN, AND MUTAGENESIS.
 RX MEDLINE=93178426; PubMed=8440242;
 RA Erdman S.E., Burtis K.C.;
 RT "The drosophila doublesex proteins share a novel zinc finger related
 RT DNA binding domain.";
 RL EMBO J. 12:527-535(1993).
 CC -!- FUNCTION: CONTROLS SOMATIC SEXUAL DIFFERENTIATION. BINDS DIRECTLY
 CC YOLK PROTEIN 1 AND 2 GENES (YPL AND YP2). THIS ENHANCER IS
 CC SUFFICIENT TO DIRECT THE FEMALE-SPECIFIC TRANSCRIPTION
 CC CHARACTERISTIC OF THE YP GENES IN ADULT FAT BODIES.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=Male;
 CC IsoId=P23023-1; Sequence=Displayed;
 CC Name=Female;
 CC IsoId=P23023-2; Sequence=VSP_001321, VSP_001322;
 CC -!- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.
 CC -!- SIMILARITY: Contains 1 DM domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M25292; AAA17840.1; -;
 CC EMBL; M25293; AAA17841.1; -;
 CC EMBL; M25294; AAA17842.1; -;
 CC EMBL; AE003676; AAF54168.1; -;
 CC PIR; A32372; A32372.
 CC PIR; B32372; B32372.
 CC PDB; 1LPV; 02-OCT-02.
 CC TRANSFAC; T00955; -;
 CC TRANSFAC; T00956; -;
 CC FlyBase; FBgn0000504; dsx.
 CC GO; GO:0003729; F:mRNA binding activity; NAS.
 CC GO; GO:0003700; F:transcription factor activity; NAS.
 CC GO; GO:0007619; P:courtship behavior; NAS.
 CC GO; GO:0045497; P:female analia morphogenesis (sensu Holometa. . ; NAS.
 CC GO; GO:0007486; P:female genital morphogenesis (sensu Holomet. . ; IGI.
 CC GO; GO:0019101; P:female somatic sex determination; NAS.
 CC GO; GO:0007483; P:genital disc metamorphosis; NAS.
 CC GO; GO:0045496; P:male analia morphogenesis (sensu Holometabola); NAS.
 CC GO; GO:0045433; P:male courtship behavior (sensu Drosophila); . ; NAS.
 CC GO; GO:0007485; P:female genital morphogenesis (sensu Holometab. . ; IGI.
 CC GO; GO:0019102; P:male somatic sex determination; NAS.
 CC GO; GO:0007548; P:sex differentiation; NAS.
 CC InterPro; IPR001275; DM_DNA-binding.
 CC Pfam; PF00751; DM-domain; 1.
 CC SMART; SM00301; DM; 1.
 CC PROSITE; PS40000; DM_DOMAIN_1; 1.

DR PROSITE; PS50809; DM DOMAIN 2; 1.
KW Sexual differentiation; Alternative splicing; DNA-binding;
KW Transcription regulation; Nuclear protein; Zinc; Metal-binding;
KW 3D-structure.
FT DNA_BIND 44 91 DM.
FT DOMAIN 119 224 HIS-RICH.
FT DOMAIN 267 296 SER/GLY-RICH.
FT VARSPLIC 398 427 ARVEINRTVAQIYYTTPMALVNGAPMYL -> GQYVVNE
FT YSRQNLNIYDGGELNRTTRQCG (in isoform
FT Female).
FT /FTIQ-VSP_001321.
FT Missing (in isoform Female).
FT /FTIQ-VSP_001322.
FT C->A,H: ABOLISHES DNA-BINDING.
FT MUTAGEN 47 47 H->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 50 50 H->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 59 59 C->D,Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 68 68 C->Y: ABOLISHES DNA-BINDING.
FT MUTAGEN 70 70 R->Q: ABOLISHES DNA-BINDING.
FT MUTAGEN 91 91
SQ SEQUENCE 549 AA; 57409 MW; 3C1B92724E4CE083 CRC64;

Query Match 3.6%; Score 146; DB 1; Length 549;
Best Local Similarity 21.0%; Pred. No. 0.0097;
Matches 57; Conservative 21; Mismatches 96; Indels 98; Gaps 8;

QY 90 IHHHDDHSDHEHSDH-----ERHSD 113
Db 129 VHAHEVHAHAGHGSHGHVHHQQAASAPASHLGGSTAASTIGHAH 188
QY 114 HEH-----HSDHEHSDHSHHHAAGKNKKAL-CPDHDSDSGKDP- 158
Db 189 HVHMAAASVACHQHQSHPHSHHHQHHPHQQPATOTALRSPPHSDHGSVC 248
QY 159 RNSQKGARPEHASGRNVKDSV-----SASEVTSTVYNTVSEGTFFETI 205
Db 249 TSSGGGAPSSNAAATSSNGSGGGGGGGGGGSGGSGGSGTSGTSGTSGTSGT 308
QY 206 ETPRPGKLPKDVSTPPSVTSKRVSLAGRTNESVSEPRKGFWSRNTNENPQECF 265
Db 309 PTPAOSLEGSCD-SSSPSPSTSGAIIPI-----SVSVNRKN----- 345
QY 266 NASKLLTSHGMGIQVPLNATEFNFLCPAINQ 297
Db 346 -----GANVPLGQDVFLDYCKLLEK 366

Search completed: September 12, 2003, 22:14:45
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2003, 21:52:06 ; Search time 88 Seconds
(without alignments)
2213.974 Million cell updates/sec

Title: US-09-642-034-5

Perfect score: 4024

Sequence: 1 MARKLSVILLITFALSVNPLHELKAAAFQTTEKISPNWESGINVDLAISTROYHLQOL 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3969	98.6	749	4 Q13433	Q13433 homo sapien
2	3542	88.0	765	11 Q8CI45	Q8CI45 mus musculu
3	2533.5	63.0	505	11 Q8R518	Q8R518 mus musculu
4	2280	56.7	433	4 Q8IXR3	Q8IXR3 homo sapien
5	2057	51.1	382	4 Q96HP5	Q96HP5 homo sapien
6	1337	33.2	835	4 Q5ULF5	Q5ULF5 homo sapien
7	902	22.4	381	11 Q8BX42	Q8BX42 mus musculu
8	787.5	19.6	701	5 Q9VSL7	Q9VSL7 drosophila
9	734	18.2	528	6 Q9GKV2	Q9GKV2 macaca fasc
10	703	17.5	228	11 Q8COL2	Q8COL2 mus musculu
11	657.5	16.3	539	4 Q8N6Y3	Q8N6Y3 homo sapien
12	652.5	16.2	535	11 Q9D856	Q9D856 mus musculu
13	652.5	16.2	535	11 Q9D909	Q9D909 mus musculu
14	640	15.9	489	11 Q8VDL0	Q8VDL0 mus musculu
15	632.5	15.7	531	4 Q15043	Q15043 homo sapien
16	616	15.3	529	4 Q9Y321	Q9Y321 homo sapien

ALIGNMENTS

RESULT 1

ID	Q13433	PRELIMINARY;	PRT;	749 AA.
AC	Q13433;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Estrogen regulated LIV-1 protein.			
GN	LIV-1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Green C., Morgan H.;			
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U41060; AAA96258.2; -			
DR	InterPro; IPR003689; zn_transpt_2ip.			
DR	Pfam; PF02535; Zip; 1.			
SQ	SEQUENCE 749 AA; 84282 NW; 6B48D4E620733F95 CRC64;			

Query Match 98.6%; Score 3969; DB 4; Length 749;

Best Local Similarity 99.2%; Pred. No. 5.8e-307; Mismatches 0; Indels 6; Gaps 1;

Matches 749; Conservative 0;

QY	1	MARKLSVILLITFALSVNPLHELKAAAFQTTEKISPNWESGINVDLAISTROYHLQOL	60
Db	1	MARKLSVILLITFALSVNPLHELKAAAFQTTEKISPNWESGINVDLAISTROYHLQOL	60
QY	61	FYRGNNLSVEGFRKLLQNIQIDTKRIHIHDDHSDHEHSHDHEHSHD 120	
Db	61	FYRGNNLSVEGFRKLLQNIQIDTKRIHIHDDHSDHEHSHDHEHSHD 120	
QY	121	EHSDHSDHSHNHAASKNKKALCPDHDSDSGKDPNSQKGAHPPEHASGRNVKD	180
Db	121	EHSDHSDHSHNHAASKNKKALCPDHDSDSGKDPNSQKGAHPPEHASGRNVKD	174
QY	181	SVSASEVTSTVNTVSEGTHLETIETPRGKLFPPKDVSSSTPPSVTSKSRVSLAGRKT	240


```

Query Match          63.0%; Score 2533.5; DB 11; Length 505;
Best Local Similarity 94.5%; Pred. No. 5.5e-193;
Matches 477; Conservative 13; Mismatches 14; Indels 1; Gaps 1;

QY 252 MYSRTNENPQECFNASKLLTSHGMIQVPLNATEFNILCPAIINQIDARSCLTHT-SEK 310
DB 1 MYSRTNDNIQECFNTKLLTSHGMSIQALINATEFNILCPAIINQIDARACLIHTASEK 60

QY 311 KAEIPPKTYLSQIANVGGFIAISIIISFLSLLGVILPLNRRVFFFLSFLVALAVGTLS 370
DB 61 KAEIPPKTYLSQIANVGGFIAISIIISFLSLLGVILPLNRRVFFFLSFLVALAVGTLS 120

QY 371 GDAFLHLLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIEESAYFDSWTKGLTALGG 430
DB 121 GDALLHLLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIEESAYFDSWTKGLTALGG 180

QY 431 LYFMFLVEHVLTLLIKQFKDKKKKKKPPENDDDVEIKKOLSKYESQSLSSTNEEKVDTDRT 490
DB 181 LYFMFLVEHVLTLLIKQFKDKKKKKKPPENDDDVEIKKOLSKYESQSLSSTNEEKVDPGERP 240

QY 491 EGYLRADSPESHFDSQOPAVLEEEVYMAHAHPQEVYNEYYPGRCKNKKCHSHFHDTLGQ 550
DB 241 EGYLRADSPESHFDSQOPAVLEEEVYMAHAHPQEVYNEYYPGRCKNKKCHSHFHDTLGQ 300

QY 551 SDDLIIHHHHYHILHHHHHQNHPHSHSORYSREELKDAGVATLAWMVGDLGHNFS 610
DB 301 SDDLIIHHHHYHILHHHHHQNHPHSHSORYSREELKDAGVATLAWMVGDLGHNFS 360

QY 611 GLAIGAAFTGSLSSGLSTSVAVFCHLPHLPGDFAVLLKAGMTVKQAVLYNALSAMLAYL 670
DB 361 GLAIGAAFTGSLSSGLSTSVAVFCHLPHLPGDFAVLLKAGMTVKQAVLYNALSAMLAYL 420

QY 671 GWATGIFIGHYAENVSMWIFALTACLEMYVALVDMPEMLHNDADSHGCSRWGFFLQNA 730
DB 421 GWATGIFIGHYAENVSMWIFALTACLEMYVALVDMPEMLHNDADSHGCSRWGFFLQNA 480

QY 731 GMLLGFGIMLLISIFEKIVFRINF 755
DB 481 GMLLGFGIMLLISIFEKIVFRINF 505

RESULT 4
Q8IXR3
ID Q8IXR3 PRELIMINARY; PRT; 433 AA.
AC Q8IXR3;
DT 01-VAR-2003 (TrEMBLrel. 23, Created)
DT 01-VAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-VAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to Liv-1 protein, estrogen regulated.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC039498; AAH39498.1; -.
SQ SEQUENCE 433 AA; 48605 MW; 72B8B90BD0A1867B CRC64;

Query Match          56.7%; Score 2280; DB 4; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.6e-173;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 MGIOVPLNATEFNILCPAIINQIDARSCLHTSEKKAIEPPKTYSLQIANVGGFIAISII 335
DB 1 MGIOVPLNATEFNILCPAIINQIDARSCLHTSEKKAIEPPKTYSLQIANVGGFIAISII 60

QY 336 SFLSLLGVILVPLMNRVFFKLLSFLVALAVGTLSGDAFLHLLPHSHASHHSHSHEEPA 395
DB 61 SFLSLLGVILVPLMNRVFFKLLSFLVALAVGTLSGDAFLHLLPHSHASHHSHSHEEPA 120

```

```

QY 396 MEMKRGPLFSLSSQNIIEESAYFDSWTKGLTALGGYFMFLVEHVLTLLIKQFKDKKKKNQ 455
DB 121 MEMKRGPLFSLSSQNIIEESAYFDSWTKGLTALGGYFMFLVEHVLTLLIKQFKDKKKKNQ 180

QY 456 KKPENDDDDVEIKKOLSKYESQSLSSTNEEKVDTDRTGELRADSQEPESHFDSQOPAVLEEE 515
DB 181 KKPENDDDDVEIKKOLSKYESQSLSSTNEEKVDTDRTGELRADSQEPESHFDSQOPAVLEEE 240

QY 516 EYMAHAHPQEVYNEYYPGRCKNKKCHSHFHDTLGQSDLLIHHHHYHILHHHHHQNHP 575
DB 241 EYMAHAHPQEVYNEYYPGRCKNKKCHSHFHDTLGQSDLLIHHHHYHILHHHHHQNHP 300

QY 576 HSHSORYSREELKDAGVATLAWMVGDLGHNFS 635
DB 301 HSHSORYSREELKDAGVATLAWMVGDLGHNFS 360

QY 636 ELPHLPGDFAVLLKAGMTVKQAVLYNALSAMLAYLGNATGIFIGHYAENVSMWIFALTAG 695
DB 361 ELPHLPGDFAVLLKAGMTVKQAVLYNALSAMLAYLGNATGIFIGHYAENVSMWIFALTAG 420

QY 696 LFMVYALVDMV 706
DB 421 LFMVYALVDMV 431

RESULT 5
Q96HP5
ID Q96HP5 PRELIMINARY; PRT; 382 AA.
AC Q96HP5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-VAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008317; AAH08317.1; -.
DR InterPro; IPR003689; Zn_transcript_2ip.
DR Pfam; PF02535; Zip; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 382 AA; 43404 MW; AE4AE1C04401CD8C CRC64;

Query Match          51.1%; Score 2057; DB 4; Length 382;
Best Local Similarity 100.0%; Pred. No. 3.1e-155;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 FLHLPLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIEESAYFDSWTKGLTALGGYLF 433
DB 1 FLHLPLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIEESAYFDSWTKGLTALGGYLF 60

QY 434 MFLVEHVLTLLIKQFKDKKKKKKPPENDDDVEIKKOLSKYESQSLSSTNEEKVDTDRTGEGY 493
DB 61 MFLVEHVLTLLIKQFKDKKKKKKPPENDDDVEIKKOLSKYESQSLSSTNEEKVDTDRTGEGY 120

QY 494 LRADSPESHFDSQOPAVLEEEVYMAHAHPQEVYNEYYPGRCKNKKCHSHFHDTLGQSD 553
DB 121 LRADSPESHFDSQOPAVLEEEVYMAHAHPQEVYNEYYPGRCKNKKCHSHFHDTLGQSD 180

QY 554 LTHHHHHYHILHHHHHQNHPHSHSORYSREELKDAGVATLAWMVGDLGHNFS 613
DB 181 LTHHHHHYHILHHHHHQNHPHSHSORYSREELKDAGVATLAWMVGDLGHNFS 240

QY 614 IGAAFTGSLSSGLSTSVAVFCHLPHLPGDFAVLLKAGMTVKQAVLYNALSAMLAYL 673
DB 241 IGAAFTGSLSSGLSTSVAVFCHLPHLPGDFAVLLKAGMTVKQAVLYNALSAMLAYL 300

```

```
QY 674 TGIFIGHYAENVSMWIFALTAGLTMVVALVDMVPEMLHNDASDHGSRMGYFFLQAGML 733
DB 301 TGIFIGHYAENVSMWIFALTAGLTMVVALVDMVPEMLHNDASDHGSRMGYFFLQAGML 360
QY 734 LGFGIMLLISIFEHKIVRINF 755
DB 361 LGFGIMLLISIFEHKIVRINF 382

RESULT 6
Q9ULF5 PRELIMINARY; PRT; 835 AA.
AC Q9ULF5;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA1265 (Fragment).
GN KIAA1265.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033091; BAA86579.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR003689; Znf_transpt_2ip.
DR Pfam; PF02355; 2ip; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 835 AA; 94559 MW; 851503EFC02F1391 CRC64;

Query Match 33.2%; Score 1337; DB 4; Length 835;
Best Local Similarity 35.5%; Pred. No. 1.8e-97;
Matches 317; Conservative 131; Mismatches 241; Indels 204; Gaps 26;

QY 1 MARKLSVILITLFTALSYTNPLHELKAAAPQTEK-----ISPWESSINVDLAISTR 53
DB 9 MTKKLCILTLTFIFHCNCHB-EHDGFEALHQHGMTELEP---SKFSKQAENEK 64
QY 54 QYHLOQLPYRGYGENNSLSVGFRLKQNLQIGDKIRIHRD---HSD----- 101
DB 65 KYVIEKLPERYGENGLSFFGLEKLLTNLGLGRKRVVEINEDLGHSHVSLDLAVQEG 124
QY 102 ---HEHSDHPRK-----SDHEHSDHSHSDH 126
DB 125 KHFFSHNHQHSHNLHNSENTVTSVSTKRHNKCDPEKETVEYSVKSDKDHMDHNLRLR 184
QY 127 DH-----HSHNHAAS-----GNKRKALCP 147
DB 185 HRLHLLHLDHNTTHFFNDSTPSERGEPSNEPTNKQSDVKLPKGRKKKGRKS 244
QY 148 DHDSO---SSGKDPNRSQK-----GAHRPEHASGRNPKDSVASEVTSTVNTVSEGH 200
DB 245 NENSEVTTPPPNHDQGEQYEHNRVHKPDR-----VNPNGSHVH 285
QY 201 FLB-----TITPPPKLFPKDPVSSSTPPSVTSKRSVRSLAGRKINESVERPK 249
DB 286 LPERNGHPDGRGHODLPDNEGEL-----RHRKREAPHVKNNAIISLRK 330
QY 250 GFMYSRNTNENPQCFNASKLLSHGMGIQVPLNATEFNILCPALINQIDARSLIHSE 309
DB 331 DL-----NEDDHHHCLNVTQLLYYGHGANSPISTDLTFLCPALLYQIDSLRCLTEHFDK 386
```

```
QY 310 KKAIEPPPKYSI-----QIAWVGGFIALSIISFLSLGVILPLMNRVFFELLSF 360
DB 387 LLVEDINKDNKLVDPEDRANTGASAWTCGLIISTVLSLSLGVILPIINQOCFFELTF 446
QY 361 LVALAVGTLSDGAPLHLLPHSHASHHSHHEEPAMEMKRGPLFSLHSQNTIEESAYFDS 420
DB 447 LVALAVGTMSGDALLHLLPHSOGGHDSHQHAH-----GHGSHGHESNKFLEE---YDA 498
QY 421 TWKGLTALGGLYFMFLVHYVLTLLIKQDKKKKQK--KPENDDDYVEIKKQSKYESQLS 478
DB 499 VLKGLVALGGIYLLFIIEHCIRMFKHKKOQKQK-KQWFMKQNTIEESTIGRKLSDKHLANT 557
QY 479 TNEEKYD-----TDDRTEGLRADSQEPSHFDSSQPA-----VLEEEVMIHAHPQEV 527
DB 558 PSDMWLQKPLAGTDDSVSEDRINETELTDLGGQESPKNYLCIEEKIIDHSDGL 617
QY 528 YNEVYPRGCKNKHSHFDFTLGSDDLIIHHHHDYHIIHLHHHHHONHHPHSHSQR---YSR 584
DB 618 HT-----IHEHDLHAAAH---NHGKNTVLRKHNHWHHKKHSHSHGPHCHSG 662
QY 585 EELKAGVATLAWVIMDGLHNFSDGLAIGAAPTEGLSSGLSTSVATCCHLPHELQDF 644
DB 663 SDLKETGIANIAWVIMDGLHNFSDGLAIGAAPTEGLSSGLSTSVATCCHLPHELQDF 722
QY 645 AVLLKAGMTVKQAVLYNALSMALAYLGMATGIFIGHYAENVSMWIFALTAGLTMVVALVD 704
DB 723 AVLLKAGMTVKQAVLYNALSMALAYLGMATGIFIGHYAENVSMWIFALTAGLTMVVALVD 782
QY 705 MYPEMLHND--SDHGSRMGYFFLQAGMLGFGIMLLISIFEHKIVRINF 755
DB 783 MLPEMLHGDGNEBHGCPVCPQTLQNLGLLFGFALMVIALYEDKIVFDIQF 835

RESULT 7
Q8BX42 PRELIMINARY; PRT; 381 AA.
AC Q8BX42;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE weakly similar to CDNA FLJ32338 FIS.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK049099; BAC33542.1; -.
SQ SEQUENCE 381 AA; 42435 MW; 9BDE487CD9765E55 CRC64;

Query Match 22.4%; Score 902; DB 11; Length 381;
Best Local Similarity 46.3%; Pred. No. 2.4e-63;
Matches 189; Conservative 67; Mismatches 104; Indels 48; Gaps 13;

QY 369 LSGDAPLHLLPHSHASHHSHHEEPAMEMKRGPLFSLHSQNTIEESAYFDSFWKGLTAL 428
DB 1 MSGDALLHLLPHSOGGHDSHQHTH-----GHGSHGHESKEFLEE---YDAVLKGLVAL 52
QY 429 GGIYFMFLVHYVLTLLIKQDKKKKQK--KPENDDDYVEIKKQSKYESQLSSTNEKVD- 485
DB 53 GGIYLLFIIEHCIRMFKHKKOQKQK-KQWFMKQSTIEESTIGRKLSDKHLNSTPDADWLQ 111
QY 486 -----TDDRTEGLRADSQEPSHFDSSQPAV-----LEEEVMIHAHPQEVNYP 534
DB 112 KPLAGTDDSVSEDRINETELTDLGAQESPKNYLVGVEEEKIM-DHSHSDGLHT----- 165
```

```

QY 535 GCKNCKCHSH-FHDTLGSODDLIHHHDYHH-ILRHHHONHH---PESHQSRYRSREELKO 589
DQ 166 -----IHHEVHVHT-----SHNHDEKAVLRKHSHQWHEHRAHSHSGPCHSSDLKE 213
QY 590 AGVATLAWMVGDLNFTSDGLAIGAFTGELSGSLTSVAVFCHLPHLGLDFAVLK 649
DQ 214 TGTIANIAWMVGDLNFTSDGLAIGAFTGELSGSLTSVAVFCHLPHLGLDFAVLK 273
QY 650 AGMTYKQAVLNALSAMAVLGMATGCTIGTGHYAENVSWKFTALTAGLFTVALVDMP 709
DQ 274 AGMTYKQAVLNALSAMAVLGMATGCTIGTGHYAENVSWKFTALTAGLFTVALVDMP 333
QY 710 LHNDL--SDHSCSWGFFYFQAGMLLGLFGLMILSTIFKHIVRINF 755
DQ 334 LHGDGDHGHGFCFVGOFILONLGLLFGFALMVLALYEDKIVDFDQF 381

RESULT 8
Q9VSL7 PRELIMINARY; PRT; 701 AA.
AC Q9VSL7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE CG6817 protein.
GN FOI OR CG6817.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
EX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moadery C., Mores J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO03555; AAF50401.2; -
DR FlyBase; FBgn0024236; foi.
DR InterPro; IPR003689; Zn_transpt.zip.
DR Pfam; PF02535; Zip; 1.
SQ SEQUENCE 701 AA; 76766 MW; 23A45DFD6C2E85F0 CRC64;

Query Match 19.68; Score 787.5; DB 5; Length 701;
Best Local Similarity 26.98; Pred No. 7.3e-54;
Matches 218; Conservative 118; Mismatches 206; Indels 267; Gaps 27;

QY 12 TFAISVTNPLHKLKAAFPQTTEKISPNWESGINVDLAISTROYHLQQLFYRTGENNSLS 71
DQ 67 TFNYSISPPSRREKRHA---GHEHGPTSESRVP-----QITQVYLEKLM---AQDELMM 114
QY 72 VSGFRKLLQNTGIDKI-----KRTIHHDHDSHDSHDSHDSHDSHDSHDSHDSHDSH 118
DQ 115 SSGFDGLLQQLSLSLASGASGEGTCVPSGRVHHVQPHDHH-----HAH-- 158
QY 119 DHEHSDHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSHSH 173
DQ 159 ---HHEEDHSLQNLNCTLLQNTGTTNVICPSLPNNHTPLGKAKNF----- 203
QY 174 GRNNKDSVSASEVTSTVYNTVSEGHFTLETITETPRPKLPKDPKDVSSSTPPSVTSKRSVS 233
DQ 204 -----TLSD-----KDLL----- 211
QY 234 RLAGRKTNEVSSEPRKGFMYSRNTNENPOCFNASKLLTSHGMGIQVPLNATEFNYLCPA 293
DQ 212 -----HLCP 216
QY 294 IINQIDABS--CL-----IHTSEKKAIPPKYTSLOIAWVGFIATISIFSLIGVI 344
DQ 217 LYLKLAGGCGCIPALISDIDITTEBLE- AEKDKDIFYWYAFISVFACGILGLVGA 275
QY 345 LVPLMNRVFFFLSFLVALAVGTLSGDAFLHLLPHSHASHHSHSHSHSHSHSHSHSHSHSHSH 404
DQ 276 IIPMGSRVYKYYIQIYLVAVGTMTGDLALLHLLPHSLA-----GQDEGIMM----- 323

```


Db 125 --PDHSGSPSLLDFQKLL--DHSIAD-----HLNE---DCLNGSQLVN 165

QY 274 HGMGIQVPLNATEFNYPALINOIDARSLIHTSEKAEIPPKYYSQIAWVGFTAIS 333

Db 166 FGLSPVAPLTPROFALLCPALLYOIDSRCI-----KTPAPAPGPDVLSALLHSLAVL 219

QY 334 IISFSLGLVILVPLMNRVFKFLLSFLVALAVGTLSGDAFLHLLPHSHASHHSHSHHEE 393

Db 220 FLSLPAPFLSLLRLGLPRLRPVGLFGLALAVGTLCGDALLHLLPHAQGGRTGTPSEQS 279

QY 394 PAMEMKRGPLFSLSSQNIIESAYFDSTWKGFLALGLYFMFLVHVLTLIKO-----446

Db 280 ---BEDLGP-----GLSVLGGLFLFMLENTLGLVHRHGLRPRC 315

QY 447 FDKKKKQKQKPPENDDDVEIKQLSKYESQLSNTEKVDTDRTGTYLRADSOBPFHDS 506

Db 316 CRNKDLGEPNPDG-----SGWLRPLQAASEVEVQGRENQSSPS----360

QY 507 QQPAVLEEEVMTAHAPQEVNYEYVPRGCKNCHSHFDTLTGQSDDLIHHHHYHILH 566

Db 361 -----LAPPG-----365

QY 567 HHHQNHHPHSHSQYSREELKDAGVATLAWMIMGDLHNFSDGLATGAATFGLSSGL 626

Db 366 -----HQHSHSHR-----GGSIAMVLLGDLNLTDLGALGAFAFSDGFSGL 409

QY 627 STSVAVFCHPELPHGDFAVLLKAGMTVKQAVLYNALSAMLAYLGMATGPIGHIYENVS 686

Db 410 STTLAVFCHPELPHGDFAMLLQBSLFRKLLLSLVSGALGGAALGVLGSLGPVPLT 469

QY 687 MTIFALTAGLFMTVALDMVPEML-----HNDASDHGCSRWGYFFLQNGMLLFGF 737

Db 470 PWVGTAGVFLYVALVDMPLTLRPPPLPVPH-----VLLQGLGLLIGGS 516

QY 738 IMLLISFEBKIV 750

Db 517 LMTFIALLEQLV 529

RESULT 13

Q9D909 PRELIMINARY; PRT; 535 AA.

AC Q9D909;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE 1810013D05RIK protein.

GN 1810013D05RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Pancreas;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Cariboldi M.,

RA Guscicich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ringwald B., Rodriquez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL; AK007473; BAB25054.1; -;

DR MGI; MGI:1919252; 1810013D05RIK.

DR InterPro; IPR003689; Zn_transpt_4ip.

DR Pfam; PF02535; Zip; 1.

SQ SEQUENCE 535 AA; 56360 MW; 77CF3310EB90DF5B CRC64;

Query Match 16.2%; Score 652.5; DB 11; Length 535;

Best Local Similarity 25.4%; Pred. No. 2.8e-43;

Matches 186; Conservative 111; Mismatches 199; Indels 237; Gaps 17;

QY 36 ISPWSEGI--NVDLAISTROYTHLQQLFYRYGNNLSVEGFRKLLQNIQDKIKRIHIH 93

Db 16 VALGWGSGVPNLGPAEQHNYLAQLFGLYGENGTITAGGLARLLHSLGLGRVQGLRLG 75

QY 94 HDHHDHSDHEHSDHERHSDHEHSDHEHSDHEHSDHSHHHAASGKNRKALCPDHSDS 153

Db 76 H-----HEPPTGRAAPTSGDNFTRL-----96

QY 154 SGKDPNRSQKGAPHEHAGSRNRVKSVSASEVTSTVNTVSEGTHTLETIETPRGKL 213

Db 97 --QEPFELSVDIWAGMPLGPGSGWQDESKA-----124

QY 214 FPKDVSSSTPPSYTSKSRVSRLAGRKTNESVSRPRKFMYSRNTNENPQECFNASKILLS 273

Db 125 --PDHSGSPSLLDFQKLL--DHSIAD-----HLNE---DCLNGSQLVN 165

QY 274 HGMGIQVPLNATEFNYPALINOIDARSLIHTSEKAEIPPKYYSQIAWVGFTAIS 333

Db 166 FGLSPVAPLTPROFALLCPALLYOIDSRCI-----KTPAPAPGPDVLSALLHSLAVL 219

QY 334 IISFSLGLVILVPLMNRVFKFLLSFLVALAVGTLSGDAFLHLLPHSHASHHSHSHHEE 393

Db 220 FLSLPAPFLSLLRLGLPRLRPVGLFGLALAVGTLCGDALLHLLPHAQGGRTGTPSEQS 279

QY 394 PAMEMKRGPLFSLSSQNIIESAYFDSTWKGFLALGLYFMFLVHVLTLIKO-----446

Db 280 ---BEDLGP-----GLSVLGGLFLFMLENTLGLVHRHGLRPRC 315

QY 447 FDKKKKQKQKPPENDDDVEIKQLSKYESQLSNTEKVDTDRTGTYLRADSOBPFHDS 506

Db 316 CRNKDLGEPNPDG-----SGWLRPLQAASEVEVQGRENQSSPS----360

QY 507 QQPAVLEEEVMTAHAPQEVNYEYVPRGCKNCHSHFDTLTGQSDDLIHHHHYHILH 566

Db 361 -----LAPPG-----365

QY 567 HHHQNHHPHSHSQYSREELKDAGVATLAWMIMGDLHNFSDGLATGAATFGLSSGL 626

Db 366 -----HQHSHSHR-----GGSIAMVLLGDLNLTDLGALGAFAFSDGFSGL 409

QY 627 STSVAVFCHPELPHGDFAVLLKAGMTVKQAVLYNALSAMLAYLGMATGPIGHIYENVS 686

Db 410 STTLAVFCHPELPHGDFAMLLQBSLFRKLLLSLVSGALGGAALGVLGSLGPVPLT 469

QY 687 MTIFALTAGLFMTVALDMVPEML-----HNDASDHGCSRWGYFFLQNGMLLFGF 737

Db 470 PWVGTAGVFLYVALVDMPLTLRPPPLPVPH-----VLLQGLGLLIGGS 516

QY 738 IMLLISFEBKIV 750

Db 517 LMTFIALLEQLV 529

RESULT 14

Q9VDLO PRELIMINARY; PRT; 489 AA.

ID Q9VDLO;

AC Q9VDLO;

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

```
DE Hypothetical 53.8 kDa protein.
GN BC021530.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021530; AAH21530.1; -.
DR MGD; MGI:2384851; BC021530.
DR InterPro; IPR003689; Zn_transpt_zip.
DR Pfam; PF02535; Zip; 1.
KW Hypothetical protein.
SQ SEQUENCE 489 AA; 53754 MW; CLE4EB928457CFC8 CRC64;

Query Match 15.9%; Score 640; DB 11; Length 489;
Best Local Similarity 29.1%; Pred. No. 2.4e-42;
Matches 166; Conservative 93; Mismatches 156; Indels 154; Gaps 18;

QY 219 SSTTPP-SVTS-----KSRVSLAGKKTNSVSEPRKGMYSRT 257
DB 30 SAGLPPLSATSELDMDRYGKNDLSLTQLKSLDLHLVGVGRDNYSQPKEG-----82

QY 258 NENPQCFNASKLITSHGMGIQVPLNATEFNILCPALINQIDARSCLIHTE--KKAETPP 316
DB 83 PRLNSCFSSGDLFAHNLSERSQIGASEFQFCFTILQQLDSQAC---TSENQKSENE 139

QY 317 KTY-----SLQIAWGGFTAIISIFSLGVLVPLMNRVFFKFLSLFVALAVGTLSD 372
DB 140 QTEEGKPSAIEVWGGFSLVSLNLASLGLVLPCTEKAFPSRVLYFYFALSIGTLN 199

QY 373 AFLHLPHSHASHHSHSHEPAMEMKRGFLPSHLSSQNIIESAYFDSWKGTLALGGLY 432
DB 200 ALFQLIPEAFGN-----PQDNIYS-----KSAVVEGGFY 229

QY 433 FMFLVHVLTILKQPKKKKNQKPPENDDDVEIKQLSKYESQLSTNEEKVDTDDRTG 492
DB 230 LFFTEKILKML-----LKQNEHHGHNH-----FTSELPKSKDOEG 269

QY 493 YLR-----ADQEPESH-----FDSQQPAVLEEEVNIHAHPQEVNEYVPRCKNKCHS 542
DB 270 VTEKLQGLDLMIPQHCSSELDGKAPGT--DEKIVNSMSVQDL-----312

QY 543 HFHDTLQSDDLLHHHHYHILHHHHQNHHPHSHSQRSREELKDAGVATLANWIMG 602
DB 313 -----QASQSAQYWLKGVYSDIGTLANWITLS 340

QY 603 DGLHNFSDGLAIGAAFTGLSSGLSTSVAVFCHPELPHGDFAVLLKAGMTVKQAVLYNA 662
DB 341 DGLHNFIDGLAIGASFTSVFQIGISTSVAILCEEPHGLDFVILLNAGMSIQQALFFNF 400

QY 663 LSAMLAYLGMATGFIG-HYAEVNSWIFALTAGLFMVYVALVDMVPEMLNDASDHGCSR 721
DB 401 LSACCCYLGLAIGLIGASHFSN---WIFALAGMFLYISLADFPPEM--NEVCQEDEN 455

QY 722 WGY---FFLQAGMLLGGIMLISIFEKI 749
DB 456 DSLVFFVIONLGLTGTSMVLVMTYSQGI 486
```

RESULT 15

```
Q15043
ID Q15043 PRELIMINARY; PRT; 531 AA.
AC Q15043;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA0062 (Fragment).
GN KIAA0062.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K., Tabata S.;
RA "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
DR EMBL; D31887; BAA06685.1; -.
DR InterPro; IPR003689; Zn_transpt_zip.
DR Pfam; PF02535; Zip; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 531 AA; 58417 MW; 5AB18EA399CCBE2B CRC64;

Query Match 15.7%; Score 632.5; DB 4; Length 531;
Best Local Similarity 30.3%; Pred. No. 1.1e-41;
Matches 159; Conservative 83; Mismatches 147; Indels 127; Gaps 16;

QY 258 NENPQCFNASKLITSHGMGIQVPLNATEFNILCPALINQIDARSCLIHTE-----SEKKA 312
DB 124 HRNLSTCFSSGDLFTAHNSEQSRIGSSSELOEFCFTILQQLDSRACTSENOENEQTE 183

QY 313 EIEPKYSLQIAWGGFTAIISIFSLGVLVPLMNRVFFKFLSLFVALAVGTLSD 372
DB 184 EGRPSAVEV---WGYGLCLVTSLSLGSAYVPMFKTFYKRLLYFYALAGTLYSN 240

QY 373 AFLHLPHSHASHHSHSHEPAMEMKRGFLPSHLSSQNIIESAYFDSWKGTLALGGLY 432
DB 241 ALFQLIPEAFG-----FNPLEDYVSKSA-----VVEGGFY 271

QY 433 FMFLVHVLTILKQPKKKKNQKPPENDDDVEIKQLSKYESQLSTNEEKVDTDDRTG 492
DB 272 LFFTEKILKIL-----LKQNEHHGH-----SHYASESLPSKK-----DQERG 311

QY 493 YLR-----ADQEPESH-----FDSQQPAVLEEEVNIHAHPQEVNEYVPRCKNKCHS 542
DB 312 VMEKLQGLDLMIPQHCSSELDGKAPW--DEKIVGSLSVQDL-----354

QY 543 HFHDTLQSDDLLHHHHYHILHHHHQNHHPHSHSQRSREELKDAGVATLANWIMG 602
DB 355 -----QASQSAQYWLKGVYSDIGTLANWITLS 382

QY 603 DGLHNFSDGLAIGAAFTGLSSGLSTSVAVFCHPELPHGDFAVLLKAGMTVKQAVLYNA 662
DB 383 DGLHNFIDGLAIGASFTSVFQIGISTSVAILCEEPHGLDFVILLNAGMSIQQALFFNF 442

QY 663 LSAMLAYLGMATGFIG-HYAEVNSWIFALTAGLFMVYVALVDMVPEMLNDASDHGCSR 721
DB 443 LSACCCYLGLAIGLIGASHFSN---WIFALAGMFLYISLADFPPEM--NEVCQED-ER 496

QY 722 WGY---FFLQAGMLLGGIMLISIFEKI 749
DB 497 KGSILPFIQNLGLTGTSMVLVMTYSQGI 528
```

Search completed: September 12, 2003, 22:16:28
Job time : 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: September 12, 2003, 22:14:11 ; Search time 29 Seconds
(without alignments)
1101.541 Million cell updates/sec
Title: US-09-642-034-5
Perfect score: 4024
Sequence: 1 MARKLSVILLITFALSVTNP.....FGIMLLISPEKIVRINF 755
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2205	54.8	431	1	US-08-311-023-2
2	388.5	9.7	417	4	US-09-599-360B-88
3	166	4.1	150	4	US-09-663-600A-196
4	154	3.8	313	3	US-08-686-528A-3
5	154	3.8	313	3	US-09-456-287-3
6	154	3.8	337	3	US-08-686-528A-2
7	154	3.8	337	3	US-09-456-287-2
8	145.5	3.6	189	1	US-08-152-922A-6
9	143.5	3.6	355	2	US-08-758-621-4
10	143.5	3.6	355	3	US-09-107-858-4
11	140	3.5	309	1	US-08-161-406-2
12	138.5	3.4	167	3	US-09-507-323B-13
13	135	3.4	466	4	US-09-134-001C-4749
14	134	3.3	126	4	US-09-663-600A-102
15	134	3.3	339	4	US-09-328-352-6551
16	132	3.3	363	4	US-09-328-352-4930
17	127.5	3.2	448	3	US-09-461-474-8
18	124.5	3.1	474	3	US-09-461-474-10
19	123	3.1	275	4	US-09-134-001C-5540
20	123	3.1	398	3	US-09-461-474-17
21	121	3.0	89	3	US-09-507-323B-12
22	121	3.0	349	3	US-09-461-474-12
23	121	3.0	649	2	US-08-149-097D-37
24	120.5	3.0	680	4	US-09-252-991A-29223
25	120	3.0	368	1	US-08-211-942-17
26	119.5	3.0	726	3	US-09-126-980-2
27	119.5	3.0	726	3	US-09-476-482-2

28	119.5	3.0	726	4	US-09-517-605-6	Sequence 6, Appli
29	118	2.9	2509	2	US-08-149-097D-35	Sequence 35, Appli
30	117.5	2.9	1182	3	US-09-041-886-21	Sequence 21, Appli
31	116	2.9	312	4	US-09-252-991A-24594	Sequence 24594, A
32	116	2.9	1261	3	US-09-208-742-4	Sequence 4, Appli
33	116	2.9	1261	4	US-09-332-295-2	Sequence 2, Appli
34	116	2.9	1261	4	US-09-709-979-2	Sequence 2, Appli
35	115.5	2.9	345	2	US-08-758-621-14	Sequence 14, Appli
36	115.5	2.9	345	3	US-09-107-858-14	Sequence 14, Appli
37	113.5	2.8	339	2	US-08-758-621-2	Sequence 2, Appli
38	113.5	2.8	339	3	US-09-107-858-2	Sequence 2, Appli
39	113	2.8	707	4	US-09-134-001C-2962	Sequence 2962, Ap
40	112.5	2.8	344	4	US-09-134-001C-3524	Sequence 3524, Ap
41	111.5	2.8	765	2	US-08-663-112-2	Sequence 2, Appli
42	111	2.8	535	2	US-08-007-107-2	Sequence 2, Appli
43	111	2.8	901	4	US-09-134-001C-5389	Sequence 5389, Ap
44	110.5	2.7	912	1	US-07-789-915A-8	Sequence 8, Appli
45	110.5	2.7	912	1	US-08-005-002C-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-311-023-2
; Sequence 2, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; TITLE OF INVENTION: BREAST TUMOURS
; NUMBER OF INVENTIONS: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32,925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-311-023-2

Query Match 54.8%; Score 2205; DB 1; Length 431;
Best Local Similarity 100.0%; Pred. No. 3e-203;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 328 GFTAIISIFSLGVLIVPLMNRVFFKLLSFLVALAVGTLSDAFLLPHSHSHH 387
DB 1 GFTAIISIFSLGVLIVPLMNRVFFKLLSFLVALAVGTLSDAFLLPHSHSHH 60

Query Match 3.6%; Score 145.5; DB 1; Length 189;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2003, 22:18:56 ; Search time 58 Seconds
(without alignments)
1899,377 Million cell updates/sec

Title: US-09-642-034-5

Perfect score: 4024
Sequence: 1 MARKUSVILITFALSVNTP.....FGIMLLISIFBHKIVPRINF 755

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	3969	98.6	749	15	US-10-177-293-264
2	3906	97.1	752	15	US-10-176-847-52
3	1972	49.0	397	10	US-09-925-300-1531
4	1337	33.2	831	9	US-09-789-561-86
5	640	15.9	1232	12	US-10-017-161-2396
6	543.5	13.5	647	10	US-09-965-529-16
7	543.5	13.5	647	11	US-09-969-680A-16
8	404	10.0	626	15	US-10-097-340-95
9	401.5	10.0	737	12	US-10-017-161-2394
10	395.5	9.8	204	15	US-10-102-806-743
11	234	5.8	354	15	US-10-050-704-115
12	234	5.8	383	15	US-10-050-704-239
13	192	4.8	292	9	US-09-864-761-37944
14	166	4.1	150	12	US-10-319-763-196
15	166	4.1	307	11	US-09-946-374-95
					Sequence 264, App
					Sequence 52, Appl
					Sequence 1531, Ap
					Sequence 86, Appl
					Sequence 2396, Ap
					Sequence 16, Appl
					Sequence 16, Appl
					Sequence 2394, App
					Sequence 743, App
					Sequence 115, App
					Sequence 115, App
					Sequence 37944, A
					Sequence 196, App
					Sequence 95, Appl

16	166	4.1	307	12	US-10-015-387A-95	Sequence 95, Appl
17	166	4.1	307	12	US-10-006-130A-95	Sequence 95, Appl
18	166	4.1	307	12	US-10-199-672-262	Sequence 262, App
19	166	4.1	307	12	US-10-006-172A-95	Sequence 95, Appl
20	166	4.1	307	12	US-10-187-749-262	Sequence 262, App
21	166	4.1	307	12	US-10-194-457-262	Sequence 262, App
22	166	4.1	307	12	US-10-184-642-262	Sequence 262, App
23	166	4.1	307	12	US-10-196-747-262	Sequence 262, App
24	166	4.1	307	12	US-10-015-392A-95	Sequence 95, Appl
25	166	4.1	307	12	US-10-017-253A-95	Sequence 95, Appl
26	166	4.1	307	12	US-10-173-689-262	Sequence 262, App
27	166	4.1	307	12	US-10-173-690-262	Sequence 262, App
28	166	4.1	307	12	US-10-173-691-262	Sequence 262, App
29	166	4.1	307	12	US-10-173-692-262	Sequence 262, App
30	166	4.1	307	12	US-10-173-694-262	Sequence 262, App
31	166	4.1	307	12	US-10-173-698-262	Sequence 262, App
32	166	4.1	307	12	US-10-173-699-262	Sequence 262, App
33	166	4.1	307	12	US-10-173-707-262	Sequence 262, App
34	166	4.1	307	12	US-10-174-569-262	Sequence 262, App
35	166	4.1	307	12	US-10-174-583-262	Sequence 262, App
36	166	4.1	307	12	US-10-174-587-262	Sequence 262, App
37	166	4.1	307	12	US-10-174-589-262	Sequence 262, App
38	166	4.1	307	12	US-10-174-591-262	Sequence 262, App
39	166	4.1	307	12	US-10-175-736-262	Sequence 262, App
40	166	4.1	307	12	US-10-175-742-262	Sequence 262, App
41	166	4.1	307	12	US-10-175-744-262	Sequence 262, App
42	166	4.1	307	12	US-10-175-745-262	Sequence 262, App
43	166	4.1	307	12	US-10-175-748-262	Sequence 262, App
44	166	4.1	307	12	US-10-175-751-262	Sequence 262, App
45	166	4.1	307	12	US-10-175-754-262	Sequence 262, App

ALIGNMENTS

RESULT 1
US-10-177-293-264
; Sequence 264, Application US/10177293
; Publication No. US20030124128A1

GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Ghatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpur, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.

TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMEN
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038

CURRENT FILING DATE: 2002-06-21

PRIOR FILING DATE: 2001-06-21

PRIOR APPLICATION NUMBER: US 60/301,572

PRIOR FILING DATE: 2001-06-27

PRIOR APPLICATION NUMBER: US 60/306,501

PRIOR FILING DATE: 2001-07-18

PRIOR APPLICATION NUMBER: US 60/325,002

PRIOR FILING DATE: 2001-09-25

PRIOR APPLICATION NUMBER: US 60/362,585

PRIOR FILING DATE: 2002-03-05

; Sequence 1531, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1531
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: SITE
; LOCATION: (179)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (180)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (181)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (358)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1531

Query Match 49.0%; Score 1972; DB 10; Length 397;
Best Local Similarity 98.7%; Pred. No. 1.5e-164;
Matches 376; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 181 SVASAEVTYVTYVSEGTHTLEIETPRPKLFPKDVSSSTPSPVTSKSRVRLAGRT 240
|||||
Db 1 SVASAEVTYVTYVSEGTHTLEIETPRPKLFPKDVSSSTPSPVTSKSRVRLAGRT 60
QY 241 NESVPRGFMYSNTNENPQECFNASKLLTSHGMGIQVPLNATFNYLCPALINQIDA 300
|||||
Db 61 NESVPRGFMYSNTNENPQECFNASKLLTSHGMGIQVPLNATFNYLCPALINQIDA 120
QY 301 RSLIHTSEKKAIEPPKTYSLQIAWVGFTAIISIFSLGLVILVPLMNRVFFKLLSF 360
|||||
Db 121 RSLIHTSEKKAIEPPKTYSLQIAWVGFTAIISIFSLGLVILVPLMNRVFFKLLXX 180
QY 361 LVALAVGTLGGDAFLHLLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIESAYFDS 420
Db 181 XVALAVGTLGGDAFLHLLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIESAYFDS 240
QY 421 TWKGLTALGGLYFMFLVEHVLTLLKQPKDKKKKNQKPPENDDDVEIKKQSKYESQLSYN 480
Db 241 TWKGLTALGGLYFMFLVEHVLTLLKQPKDKKKKNQKPPENDDDVEIKKQSKYESQLSYN 300
QY 481 EEKVYDTRTEGYLRADSQSPSHFDSQQPAVLEEEVMTAHAAHPOEVYNEYVPRGCKNC 540
Db 301 EEKVYDTRTEGYLRADSQSPSHFDSQQPAVLEEEVMTAHAAHPOEVYNEYVPRGCKXC 360
QY 541 HSHFHDPLGQSDLLHHHDY 561
Db 361 HSHFHDPLGQSDLLHHHDY 381

RESULT 4

US-09-789-561-86
; Sequence 86, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.

; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: PZ043F1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-789-561-86

Query Match 33.2%; Score 1337; DB 9; Length 831;
Best Local Similarity 35.5%; Pred. No. 2.7e-108;
Matches 317; Conservative 131; Mismatches 241; Indels 204; Gaps 26;

QY 1 MARKLSVILLIHLFALSVTNPLHKLKAAAFPTTEK-----ISPWESGINDVLAISRP 53
Db 5 MHTKFLICLLIFIPHHNCHCHE-EHDHGPALHROHGMTELEP---SKFSQAAENK 60
QY 54 QYHLOQLFYRYGENSLSVGEGRKLLQNIQIDKIRIHHD---HDI--HSD----- 101
Db 61 KYVTEKLFERYGNGRLFFGLEKLLNLGLGRKVVIEINHDGLGHVSHLDILAVQBG 120
QY 102 ---HEHSDHERH-----SDHEHSDHEHSDH 126
Db 121 KHFSHNHGHSHNHLNSNQVTSVTKRNHKODEKETVSVSKSDKHMDHNRHLRH 180
QY 127 DH-----HSHNHAAS-----GKNKRKALCP 147
Db 181 HRLHLLHLDHNTTHFHNDSTTPSGEPSPNEPSTETNKTOEQSDVKLPKGRKKGRKS 240
QY 148 DHUSD--SSGKOPRNSQK-----GAHRPEHAGRRNWKDSVSASEVTSYVYVSEGYH 200
Db 241 NENSEVITRGFPFNHDOGEQYEHNRVHKDPR-----VHNPFGSHVH 281
QY 201 FLE-----TIETPRPKLFPKDVSSSTPSPVTSKSRVSLAGRTNESVSEPRK 249
Db 282 LPERNGHDPGRGHQDLDPDNEGEL-----RHTRKEAPHVKNNAILSLRK 326
QY 250 GFMYSRNTNENPQECFNASKLLTSHGMGIQVPLNATFNYLCPALINQIDARSCLIHSE 309
Db 327 DL---NEDDHHHECLNTVQLKYVGHGANSPISTDLFTYLCALLYIDSRLCIEHFDR 382
QY 310 KKAIEPPKYSL-----QIAWVGFTAIISIFSLGLVILVPLMNRVFFKLLSF 360
Db 383 LLVEDINKNLVPEDEANIGASAWICGIIISITVLSLLGLVILVPLINGOCFKLLTF 442
QY 361 LVALAVGTLGGDAFLHLLPHSHASHHSHSHEEPAMEMKRGPLFSLSSQNIIESAYFDS 420
Db 443 LVALAVGTLGGDAFLHLLPHSQGGHDHSHQHAH-----GHGSHSHESNKYLEE---YDA 494
QY 421 TWKGLTALGGLYFMFLVEHVLTLLKQPKDKKKKNQK--KPENDDVEIKKQSKYESQLS 478
Db 495 VLKGLVALGGLYLLFTIEHCIRMFKHYKQQRGK-QKWFMKQNTSESTIGRKLSDHKLNT 553
QY 479 TNEEKVD-----TDDRTEGYLRADSQSPSHFDSQQPA-----VLEEEVMTAHAAHPOEV 527
Db 554 PSDWLQLKPLAGTDSDSVVSEDRNLNETELDLEGQOQSPPKNYLCIEBEKIIDHSHDGL 613
QY 528 YNEVYPRGCKNCKSHFHDTLGQSDLLHHHDYHILHLLHHHONHHHPHSHSQR---YSR 584
Db 614 HT-----IHEHDLHAAAH---NHHGENKTVTLKKNHNNHWHHKKHSHSGHSGCHSG 658
QY 585 EEKLDAGVATLAWVIMGDIHNFSDGLAIGAAGFTBGLSGSLSTSVAVFCHLPHELJGDF 644
Db 659 SDLKETGIANIWWIMGDIHNFSDGLAIGAAGFTBGLSGSLSTSVAVFCHLPHELJGDF 718

QY 645 AVLLKAGTQVQAVLYNALSMAYLGMATGIFIGHYAENVSMWIFALTAGLFTVALVD 704
Db 719 AVLLKAGTQVQAVLYNALSMAYLGMATGIFIGHYAENVSMWIFALTAGLFTVALVD 778
QY 705 MYPEMLHDA--SDHGSRWGYFFLQNAAGMLGFGIMLLISIFEHIVFRINF 755
Db 779 MLPEMLHGDGNEBHGFCVPVQFILQNLGLLFGFALMIVIALYEDKIVEDIQF 831

RESULT 5
US-10-017-161-2396
; Sequence 2396, Application US/10017161
; Publication NO. US2003014368A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-2396

Query Match 15.9%; Score 640; DB 12; Length 1232;
Best Local Similarity 26.1%; Pred. No. 8e-47;
Matches 189; Conservative 107; Mismatches 181; Indels 246; Gaps 22;
QY 40 WESGI--NVDLAISTROYHLQQLFYRYGENNLSVEGFRKLQNLQIGDKIKRIHHDDH 97
Db 20 WVGSGVPLNGLPAEQEONHYLAQLFGLYGENGTLAGGLARLLHSLGLGRVQGLRL---- 74
QY 98 HHSDEHSDHERSHDEHSHDHDHSHHSHHAAAGKNNKRALCPDHDSDSSGKD 157
Db 75 -----HRQN-----PELSVDVWAGMPLGPGWGDLEESKAPH-----L 127
QY 158 PRNSOGKAHPEHASGRNRVKSASVSTVYNTV---SEGTHELETETPRPKLF 214
Db 94 -----HRQN-----PELSVDVWAGMPLGPGWGDLEESKAPH-----L 127
QY 215 PKDVSSPTSPVTSKSRVSLAGRKTNESVSEPRKGFMYSRNTNENPQCFNASKLLTSH 274
Db 128 PRG-----PAPSGLDLHLRL--LLDHSAD-----HLNE--DCLNGSOLLVNF 167
QY 275 GMGIQVPLNATFEYNLCPALINOIDARSCLHTSEKKAETPP---KTSLOIAVWGGFI 330
Db 168 GLSPAAPTTPQFALCPALLXQIDSRVCI---GAPAPAPGDLSSGGQGVALLQSAL 223
QY 331 AITSIFLSLGLVILPLMNRVFFKLLSFLVALAVGTLSDAFLHLLPHSHASHSHS 390
Db 224 AVLLSLSPISLILLRLPLRLPLGLFGLALAVGLCGDALLHLLPHAGEGR---HA 280
QY 391 HEPAMEMKRGPLFSLSSONIBESAYFDSTWKGTLGGLGYFMFLVEHLVLIKQFDK 450
Db 281 GPGLPEKDLGP-----GLSVLGLFLFVLENNMLGLRH--- 315
QY 451 KKNOKKPENDDEVEIKKQSKYESOLSTNEEKVDTDRREGYLRADSPHSDQOQA 510
Db 316 RGLRPVKP-----GAQOREKNSQHP-----PA 338
QY 511 VLEEEVMTAAHPQEVYNEVPRGCKNCKSHSHFDTLGQSDLLIHHHDYHLLHHH 570
Db 339 L-----APPG----- 343

QY 571 QNEHPHSHSQYSREELKDAGVATLAWMIVMGDLHNEFSDGLAIGAAFTGELSSGLSTSV 630
Db 344 --HGHSHGHQ-----GGTDIWMVLLGDGLHNTDGLAIGRAFSDFSGSLSTL 392
QY 631 AVFCHLPHLGDPAVLKAGMTVQAVLYNALSMAYLGMATGIPIGHYAENVSMWIF 690
Db 393 AVFCHLPHLGDPAVLKAGMTVQAVLYNALSMAYLGMATGIPIGHYAENVSMWIF 690
QY 691 ALTAGLFWVALVDMVPEMLHND--ASDHGSRWGYFFLQNAAGMLGFGIMLLISIFEH 747
Db 453 GVTAGVFLYVALVDMPLPRLPPEPLTPH-----VLLQGLGLLGLGGLMIAITLLEE 505
QY 748 KIV 750
Db 506 RLL 508

RESULT 6
US-09-965-529-16
; Sequence 16, Application US/09965529
; Publication NO. US20020182671A1
; GENERAL INFORMATION:
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LO, Dying Aina M.
; APPLICANT: PATTERSON, Chandra
; TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0731 USA
; CURRENT APPLICATION NUMBER: US/09/965,529
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/149,641; 60/164,203; PCT/US00/22315
; PRIOR FILING DATE: 1999-08-17; 1999-11-09; 2000-08-14
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PERL Program
; SEQ ID NO 16
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020182671A1 1336728CD1
US-09-965-529-16

Query Match 13.5%; Score 543.5; DB 10; Length 647;
Best Local Similarity 26.5%; Pred. No. 8.9e-39;
Matches 178; Conservative 99; Mismatches 201; Indels 193; Gaps 23;
QY 130 SHNHAASGKNNKRALCPD-----HDSGSGKDPNRS-----QCKGAHREHAS 173
Db 121 SHADHLLALLSPALKTPGLSWLLQRMQARAAGQPKTACVDIPQLLEAVAGAGPAGSAG 180
QY 174 GRNVKDSVSASEVTYNTVSEGTHFETETPRPKLFPKDVSSPTSPVTSK--SR 231
Db 181 G-----VLAALLHVRSGSCF--HALPSPQYFVDFVFOOHSSEVPMTLBSAL 227
QY 232 VSKL--AGRKTNESVSEPRKG-----FMYSRNTNENPQCFNASKLLTSHGMGIQVP 281
Db 228 MQLGVGREAHSDSHRHGASSRDVPVLISSSNSSSVMDTVCLSRDYMVAAGVGLSEQAG 287
QY 282 LNATFENILCPALINOIDARSCLHTSEKKAETPPKTSLOIA--WVGFIATISIFLS 339
Db 288 VTPEAWAQLSPALLQQQLSGAC---TSQSR---PPVQDLSQSERVLYGSLATLILCLA 341
QY 340 LLGVILVPLMN--RYFEKFLLSFLVALAVGTLSDAFLHLLPHSHASHSHSHEEPAMEM 398
Db 342 VFLGLTCTGCGVTHYIIQTFLSLAVGALTGDVHLHTPKVLGLHTHS----- 391
QY 399 KRGLPLFSLSSONIBESAYFDSTWKGTLGGLGYFMFLVEHLVLIKQFDKKNQKPK 458

; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-95

Query Match 10.0%; Score 404; DB 15; Length 626;
Best Local Similarity 24.7%; Pred. No. 1.5e-26;
Matches 146; Conservative 86; Mismatches 173; Indels 186; Gaps 20;
QY 130 SHHNAASGKRRKALCPD-----HDSGSGKDRPNS-----QKGGAHRPHAS 173
DB 96 SHADHLLALLESKALTPLGSLWLLQRMQARAAGTPTACVDIPQLLEAEVAGAGAGSAG 155
QY 174 GRNVKUSVASEVTNTVTSGETHFEIETPRPKLFPKDVSSSTPPSVTSK--SR 231
DB 156 G-----VLAALLDHRVSGSCF-HALPSPQYFVDFVQQHSSEVPMTLAELSA 202
QY 232 VSRL-AGRTNESVSPRKG-----FMSNTNENPQECFNASKLLTSHGMIQVP 281
DB 203 MRLGVGREASHSHRGASSRDPVPLISSNSSSSWDTVCLSDARDVMAAYGLSQAG 262
QY 282 LNATEFNYPALINQIDARSCLHTSEKKAIEIPPKTYSIQIA--WVGFTAIISIIFLS 339
DB 263 VTPEAWAQLSPALLQO---QLSGAYTSQSR---PPVQDLSQSERVLYGLSLATLLIC 316
QY 340 LIGVLYPLMN-RVFFKFLSLFVALAVGTLGDAFLHLLPHSHASHHSHSHEEPAMEM 398
DB 317 VEGLLLTCTCGRGVAHYLIQTLFLAVGALTGDVAVLHLPKYLGLHTHS----- 366
QY 399 KRGPLFSLSSQNTIESAYFDSTWKGTLALGLYFMFLVHLLTIKQFKDKKKNQKPP 458
DB 367 -----EEGLSPQPTWRLAMLAGLYAFLFENLNL-----LP 400
QY 459 ENDDVVE-----TKKOLSKYESQSLTNEEKVDNDRTEGVLRADSOEP 501
DB 401 RDPEDLEDCGCHSHSHGHSHGVSQLPASE-----LR-QPKP 439
QY 502 SHEDSQOPAVLEEEVEVMAHAHPQEVYNEVYPRGCKNKHSHPHDPLGOSDDLHHHHDY 561
DB 440 PHEGSRADLVAE-----SPELL----- 457
QY 562 HHILHHHHQNHHPHSHSQRYSEELKDGAVATLAWVIMGDLGNHNSDGLAIGAFTGEG 621
DB 458 -----NPEP-----LRLSPF-----LRLPYMITLGDVHNFADGLAVGAFASS 497
QY 622 LSSGLSTSVAVFCHPELHFGDAFVILLKAGTVYKQAVLYNALSAMLAYLGM 672
DB 498 WKTGLATSLAVFCHPELHFGDAFVILLKAGTVYKQAVLYNALSAMLAYLGM 548

RESULT 9

US-10-017-161-2394
; Sequence 2394, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789

; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2394
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-2394

Query Match 10.0%; Score 401.5; DB 12; Length 737;
Best Local Similarity 23.3%; Pred. No. 3.2e-26;
Matches 165; Conservative 95; Mismatches 205; Indels 243; Gaps 25;
QY 130 SHHNAASGKRRKALCPD-----HDSGSGKDRPNS-----QKGGAHRPHAS 173
DB 121 SHADHLLALLESKALTPLGSLWLLQRMQARAAGTPTACVDIPQLLEAEVAGAGAGSAG 180
QY 174 GRNVKUSVASEVTNTVTSGETHFEIETPRPKLFP-KDYSSSTPPSVTSKRV 232
DB 181 G-----VLAALLDHRVSGSCF-HALPSPQYFVDFVQQHSSEVPMTLAASL 227
QY 233 SLACRKTNE-----SVSEPRKGFMYSRNTN 258
DB 228 GGRAGHTDRCQLFSLPELSALMQLGVGREASHSHRGASSRDPVPLISSNSSS 287
QY 259 ENPQECFNASKLLTSHGMIQVPLNATEFNLYCPALINQIDARSCLHTSEKKAIEIPPKT 318
DB 288 VMDIVCLSDARDVMAAYGLSEQAGVTPEAWAQLSPALLQOOLSGAC---TSQSR---PPVQ 341
QY 319 YSLQIA--WVGFTAIISIIFLSLGLVILPLMN-RVFFKFLSLFVALAVGTLGDAFL 375
DB 342 DOLQSERYLYGLSLATLLICLCAVFLGILLTCTCGRGVTHYLIQTLFLSLAVGAVTGDAVL 401
QY 376 HLLPHSHASHHSHSHEEPAMEMKRGPLFS-----HLSQNTIESAYFDSTWKG--L 425
DB 402 HLP-----KVCPRKATLALSPWPPTPCWGCIIHTAKRASASP-----PGALL 447
QY 426 TALGLYFMFLVHLLTIKQFKDKKKNQKPPENDDDVEIKKOLSKYESQSLTNEEKVD 485
DB 448 AMLAGLYAFLFENL-----NLQPRDPEDLE----- 475
QY 486 TDRTEGVLRADSOEPFSDHDSQOPAVLEEEVEVMAHAHPQEVYNEVYPRGCKNKHSHF 545
DB 476 --DGPCGCHSHSHGHSHGVSQLA-----PSELQPKPP-----H 509
QY 546 DPLGOSDDLHHHHDYHHILHHHONHHP-HSHSOR-----VSREELK 588
DB 510 E--GSRADLV-----SGRQMPHPTFSHPRLPTPLPAPQACGSARE--- 552
QY 589 DAGVATLAWVIMGDLGNHNSDGLAIGAFTGEGSLGSLSTSVAVFCHPELHPEL----- 641
DB 553 ---LRLPYMITLGDVHNFADGLAVGAFAFAFWKTGLATSLAVFCHPELHGERRGL 609
QY 642 -----GDFAVLLKAGTVYKQAV 658
DB 610 EDGGRGTGLGGETWNNKAGPRWAGPHLAGWDRKQVGRGLEGDFALHRLAGLSVROL 669
QY 659 LYNALSAMLAYIGMATGIFIGHYAEVNSVMWIFALTAGIFMYVALVDMV 706
DB 670 LLNLASALTAFAGLYVALAVG-VSESEAWIATVATGIFLYVALCDMV 716

RESULT 10

US-10-102-806-743
; Sequence 743, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298

```

; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCV/JUS00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0

```

Query Match 9.8%; Score 395.5; DB 15; Length 204;
Best Local Similarity 49.4%; Pred. No. 1.5e-26;
Matches 78; Conservative 29; Mismatches 44; Indels 7; Gaps 3;

QY 714 ASDHGRWGTYFQLQAGMLLGFIMLLISIFEHKIVF 751
 | | | | | | | | | | | | | | | | | | | | | |
 Db 172 PRP-----WLLFLHNVLGGTWTLLLSLYEDDITF 204

```

RESULT 11
US/10-050-704-115
/ Sequence 115, Application US/10050704
/ Publication No. US20030050442A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruben et al.
/ TITLE OF INVENTION: 62 Human Secreted Proteins
/ FILE REFERENCE: P2039P1
/ CURRENT APPLICATION NUMBER: US/10/050,704
/ CURRENT FILING DATE: 2002-01-18
/ PRIOR APPLICATION NUMBER: 09/684,524
/ PRIOR FILING DATE: 2000-10-10
/ PRIOR APPLICATION NUMBER: PCF/US00/08979
/ PRIOR FILING DATE: 2000-04-06
/ PRIOR APPLICATION NUMBER: 60/128,693
/ PRIOR FILING DATE: 1999-04-09
/ PRIOR APPLICATION NUMBER: 60/130,991
/ PRIOR FILING DATE: 1999-04-26
/ NUMBER OF SEQ ID NOS: 344
/ SOFTWARE: Patent Ver. 2.0

```

Query Match 5.8%; Score 234; DB 15; Length 354;
Best Local Similarity 20.4%; Pred. No. 5.4e-12;
Matches 96; Conservative 71; Mismatches 145; Indels 158; Gaps 14;

85	Db	85	TMLSEAGAWRLKQLLSFALGGLGNLGNFHLPEAWA-----YTC	124
407	QY	407	LSSONIEZSAYFDSTWKGITALGGLYFMFLVHVLILIKQFDKDKKKKQKXPENDDVEI	466
125	Db	125	SASPGGGQLQQOQQLGLWVIAGTLTF-----LALKMFLDSKEGTQAPNKDPTAA	178
467	QY	467	KKQLSKYESQLSINESKVDTDRTGYILRADSQEPSHFDSSQPAVLVEEEVWIAHAHPQE	526
179	Db	179	AAALN-----GGHCLA-----QPA-----AEP--	195
527	QY	527	VIYEVVPRGCKNCKSHFHDTLGQSDDLIHHHDYHHILLHHHHQNHHPHSHSQRSREE	586
196	Db	196	-----GLGA-----VVRS	203
587	QY	587	LKDAGVATLAWMVGIDGLHNPESDGLATGAETGLSSGLSTSVAFPCHELPHGLGDFAV	646
204	Db	204	IKVSG-----YLMNLANTIDNETHGLAVASFLVSKYKIGLLTMAILHEHPHVGDFAI	258
647	QY	647	LKAGMTVKQAVLYNALMSALMAYLGMATGIFIG-----HYAENVSMWIFALTAGL	696
259	Db	259	LIRAG-----FDRWSAAKLQLSTALGGLGAGFAICTQSPKGVGEETAAWLPTISGG	310
697	QY	697	FWTVALVDNTPMELHNDASHGCSKNGYFFLQNGMI-LGFGIMLLISIF	745
311	Db	311	FLYLLANVLPDLLEEE-----DPWR-----SLQQLLLICAGIVMWLFSLF	352

RESULT 12

```

US-10-050-704-239
> Sequence 239, Application US/10050704
> Publication No. US20030050442A1
> GENERAL INFORMATION:
> APPLICANT: Ruben et al.
> TITLE OF INVENTION: 62 Human Secreted Proteins
> FILE REFERENCE: PZ039P1
> CURRENT APPLICATION NUMBER: US/10/050,704
> CURRENT FILING DATE: 2002-01-18
> PRIOR APPLICATION NUMBER: 09/684,524
> PRIOR FILING DATE: 2000-10-10
> PRIOR APPLICATION NUMBER: PCT/US00/08979
> PRIOR FILING DATE: 2000-04-06
> PRIOR APPLICATION NUMBER: 60/128,693
> PRIOR FILING DATE: 1999-04-09
> PRIOR APPLICATION NUMBER: 60/130,991
> PRIOR FILING DATE: 1999-04-26
> NUMBER OF SEQ ID NOS: 344
> SOFTWARE: Patentin Ver. 2.0
> SEQ ID NO 239
> LENGTH: 383
> TYPE: PRT
> ORGANISM: Homo sapiens
> FEATURE:
> NAME/KEY: SITE
> LOCATION: (39)
> OTHER INFORMATION: xaa equals any of the natu
US-10-050-704-239

```

[illegible]


```

; PRIOR APPLICATION NUMBER: 60/104257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match          4.1%; Score 166; DB 11; Length 307;
Best Local Similarity 24.8%; Pred. No. 4.1e-06;
Matches 68; Conservative 40; Mismatches 80; Indels 86; Gaps 12

QY 532 VPRCKNCKHSFHDTLQSGDDLLHHHDYHH-----ILHHHHQHHPHSH- 578
   || : : : : : ||
Db 55 VPEG---VHALYEDILEKK---HQAQETHNVNIASDKAAEKSVMHEHSHDHTQLHA 106

QY 579 -----SQRYSEEEKLDAGVATLANWYMGDGLHNFSDGLAIG-A 616
   || : : : : : ||
Db 107 YIGVSLVGFVPMLLVDQIGNSHSVHSTDDEFAARSSNKITTLGLVVHAAADGVALGAA 166

QY 617 AFTBGLSSGLTSVAVFCHPELPHSGDFAVLLKAGM---TVKQAVLYNALSA---MLAY 669
   || : : : : : ||
Db 167 ASTSQTSVQLIVFVALMLHKAPAAFGVLSELMHAGLERNRINKHLLVFALAAPVMSWVY 226

QY 670 LGM-----ATGICTGHYAENVSMWIEPALTAGLFTMYVALVDKVPB---MLHNDAA 714
   || : : : : : ||
Db 227 LGLSKSKSEALSEVNATGV-----AMLFSGAGTFLYVATVHYLPEVGGIGHSHK 274

QY 715 SD-----HGCRCWGYFFLQNMAGMLLGFIMLLISI 744
   || : : : : : ||
Db 275 PDATGGGRLSR-----LEVAALVLGCLPLILSV 303

Search completed: September 12, 2003, 22:27:24
Job time : 65 secs

```